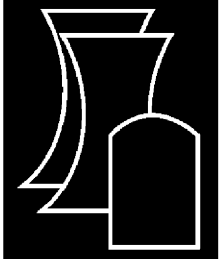
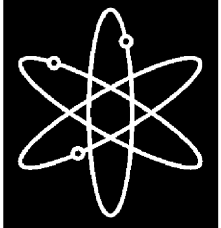
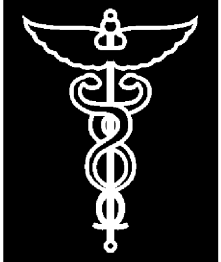
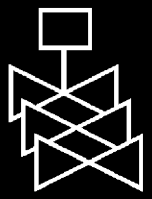
NUREG/CR-6823
SAND2003-3348P









Handbook of Parameter Estimation for Probabilistic Risk Assessment

Sandia National Laboratories

U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research
Washington, DC 20555-0001

NUREG/CR-6823
SAND2003-3348P

Handbook of Parameter Estimation for Probabilistic Risk Assessment

Manuscript Completed: November 2002
Date Published: September 2003

Prepared by
C.L. Atwood[1], J.L. LaChance[2], H.F. Martz[3],
D.J. Anderson[2], M. Englehardt[4], D. Whitehead[2],
T. Wheeler[2]

[2]Sandia National Laboratories
Albuquerque, NM 87185-0748

[1]Statwood Consulting
Silver Spring, MD 20910

[3]Los Alamos National Laboratory
Los Alamos, NM 87545

[4]Formerly with Idaho National Engineering and Environmental Laboratory
Idaho Falls, ID 83415

A.D. Salomon, NRC Technical Monitor

Prepared for
Division of Risk Analysis and Applications
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001
NRC Job Code W6970

ABSTRACT

Probabilistic risk assessment (PRA) is a mature technology that can provide a quantitative assessment of the risk from accidents in nuclear power plants. It involves the development of models that delineate the response of systems and operators to accident initiating events. Additional models are generated to identify the component failure modes required to cause the accident mitigating systems to fail. Each component failure mode is represented as an individual "basic event" in the systems models. Estimates of risk are obtained by propagating the uncertainty distributions for each of the parameters through the PRA models.

The data analysis portion of a nuclear power plant PRA provides estimates of the parameters used to determine the frequencies and probabilities of the various events modeled in a PRA. This handbook provides guidance on sources of information and methods for estimating the parameters used in PRA models and for quantifying the uncertainties in the estimates. This includes determination of both plant-specific and generic estimates for initiating event frequencies, component failure rates and unavailabilities, and equipment non-recovery probabilities.

Table of Contents

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

LIST OF TABLES

LIST OF EXAMPLES

FOREWORD

During the last several years, both the U.S. Nuclear Regulatory Commission (NRC) and the nuclear industry have recognized that probabilistic risk assessment (PRA) has evolved to the point where it can be used in a variety of applications including as a tool in the regulatory decision-making process. The increased use of PRA has led to the conclusion that the PRA scope and model must be commensurate with the applications. Several procedural guides and standards have been and are being developed that identify requirements for the PRA models. For example, the "Standard For Probabilistic Risk Assessment For Nuclear Power Plant Applications" published by The American Society of Mechanical Engineers (ASME) in 2002 (ASME-RA-S-2002) defines requirements for PRA analysis used to develop risk-informed decisions for commercial nuclear power plants, and describes a process for applying these requirements in specific applications. This handbook was generated to support these documents by providing a compendium of good practices that a PRA analyst can use to generate the parameter distributions required for quantifying PRA models.

The increased use of risk assessment has also helped promote the idea that the collection and analysis of event data is an important activity in and of itself. In particular, the monitoring of equipment performance and evaluation of equipment trends can be used to enhance plant performance and reliability. The reference material provided in this handbook can support those efforts.

This handbook provides references on sources of information and methods for estimating parameter distributions. This includes determination of both plant-specific and generic estimates for initiating event frequencies, component failure rates and unavailability, and equipment non-recovery probabilities, all of which directly supplement the ASME PRA standard.

This handbook provides the basic information needed to generate estimates of the parameters listed above. It begins by describing the probability models and plant data used to evaluate each of the parameters. Possible sources for the plant data are identified and guidance on the collection, screening, and interpretation is provided. The statistical techniques (both Bayesian and classical methods) required to analyze the collected data and test the validity of statistical models are described. Examples are provided to help the PRA analyst utilize the different techniques.

This handbook also provides advanced techniques that address modeling of time trends. Methods for combining data from a number of similar, but not identical, sources are also provided. This includes empirical and hierarchical Bayesian approaches. Again, examples are provided to guide the analyst.

This handbook does not provide guidance on parameter estimation for all types of events included in a PRA. Specifically, common cause failure and human error probabilities are not addressed. In addition, guidance is not provided with regard to the use of expert elicitation. For analysis of these events, the PRA analyst should consult other sources, some of which are cited in Chapter 1.

ACKNOWLEDGMENTS

The authors wish to acknowledge Dennis C. Bley, Iyiin Chang, Mary T. Drouin, Leslie E. Lancaster, Gareth W. Parry, Dale M. Rasmuson, David Robinson, Arthur D. Salomon, and Nathan O. Siu, who gave extensive and insightful written comments. In particular, it was Mary Drouin's initiative that started the project. She provided the early guidance and support for the development of the document. Bley and Parry not only provided comments, but also contributed text that was included in the final report. W. Jay Conover gave helpful advice on tests of serial independence, and David S. Moore gave advice on the relation of the chi-squared statistic and the deviance. Robert F. Cavedo and Steven A. Eide supplied information on the treatment of unavailability in current PRAs. Robert W. Youngblood contributed ideas and references on the relation of the two models for failure to start. Steven M. Alferink, William J. Galyean, Cynthia D. Gentillon, Dana L. Kelly, and Martin B. Sattison supplied examples and constructive comments. Ralph Nyman provided information about the Swedish I-Book. Teresa Sype provided help in formatting the report. Christine E. White produced many of the figures.

ABBREVIATIONS

AC	alternating current
AFW	auxiliary feedwater
ALWR	Advanced Light Water Reactor
ANO	Arkansas Nuclear One
ANSI	American National Standards Institute
ASEP	Accident Sequence Evaluation Program
ASME	American Society of Mechanical Engineers
ATWS	anticipated transient without scram
BWR	boiling water reactor
CDF	core damage frequency
c.d.f.	cumulative distribution function
ComEd	Commonwealth Edison
ConEd	Consolidated Edison
CPC	Consumers Power Company
CREDO	Centralized Reliability Data Organization
CVC	chemical and volume control
DC	direct current
DG	diesel generator
DOE	Department of Energy
ECCS	emergency core cooling system
EDF	empirical distribution function
EDG	emergency diesel generator
EEI	Edison Electric Institute
EF	error factor
EPIX	Equipment Performance and Information Exchange System
EPRI	Electric Power Research Institute
FTR	failure to run
FTS	failure to start
GE	General Electric
HCF	Hardware Component Failure
HEP	human error probability
HHRAG	Human and Hardware Reliability Analysis Group
HPCI	high pressure coolant injection
HPP	homogeneous Poisson process
HRA	human reliability analysis
HVAC	heating, ventilation and air conditioning
IEEE	Institute of Electrical and Electronics Engineers
IGSCC	intergranular stress corrosion cracking
i.i.d.	independent identically distributed
INEL	Idaho National Engineering Laboratory
INEEL	Idaho National Engineering and Environmental Laboratory
INPO	Institute of Nuclear Power Operations
IPE	Individual Plant Examination
IPEEE	Individual Plant Examination of External Events
IPRDS	In-Plant Reliability Data Base for Nuclear Power Plant Components
IREP	Interim Reliability Evaluation Program
LER	Licensee Event Report
LCO	limiting conditions of operation
LOCA	loss of coolant accident
LOSP	loss of offsite power
MCMC	Markov Chain Monte Carlo

MLE	maximum likelihood estimate, or maximum likelihood estimator, depending on the context (see *estimate* in glossary)
MOV	motor-operated valve
MSE	mean square error
MSIV	main steam isolation valve
MTTF	mean time to failure
MTTR	mean time to repair
NA	not applicable
NHPP	nonhomogeneous Poisson process
NPE	Nuclear Power Experience
NPP	Nuclear Power Plant
NPRDS	Nuclear Plant Reliability Data System
NRC	U.S. Nuclear Regulatory Commission
NREP	Nuclear Reliability Evaluation Program
NSAC	Nuclear Safety Analysis Center
NUCLARR	Nuclear Computerized Library for Assessing Reactor Reliability
OREDA	Offshore Reliability Data
PCS	Power Conversion System
PECO	Philadelphia Electric Company
p.d.f.	probability density or distribution function
PLG	Pickard, Lowe and Garrick, Inc.
PORV	power operated relief valve
PRA	Probabilistic Risk Assessment
PSA	Probabilistic Safety Assessment
PWR	pressurized water reactor
RAC	Reliability Analysis Center
RADS	Reliability and Availability Data System
RCIC	reactor core isolation cooling
RHR	residual heat removal
RMIEP	Risk Methods Integration and Evaluation Program
RPS	reactor protection system
RSSMAP	Reactor Safety Study Methodology Application Program
SAIC	Scientific Applications International Corporation
SBLOCAs	small-break loss-of-coolant accidents
SCSS	Sequence Coding and Search System
SKi	Swedish Nuclear Power Inspectorate
SRV	safety relief valve
SSPI	Safety System Performance Indicator
SwRI	Southwest Research Institute
U.K.	United Kingdom
U.S.	United States

1. INTRODUCTION

1.1 Objective

The data analysis portion of a nuclear power plant probabilistic risk assessment (PRA) provides estimates of the parameters used to determine the frequencies and probabilities of the various events modeled in a PRA. The objective of this handbook is to provide methods for estimating the parameters used in PRA models and for quantifying the uncertainties in the estimates.

1.2 Background

Probabilistic risk assessment is a mature technology that can provide a quantitative assessment of the risk from accidents in nuclear power plants. It involves the development of models that delineate the response of systems and operators to accident initiating events. Additional models are generated to identify the component failure modes required to cause the accident mitigating systems to fail. Each component failure mode is represented as an individual "basic event" in the systems models. Estimates of risk are obtained by propagating the uncertainty distributions for each of the parameters through the PRA models.

During the last several years, both the U.S. Nuclear Regulatory Commission (NRC) and the nuclear industry have recognized that PRA has evolved to the point where it can be used in a variety of applications, including as a tool in the regulatory decision-making process. The increased use of PRA has led to the conclusion that the PRA scope and model must be commensurate with the applications. Several procedural guides and standards have been and are being developed that identify requirements for the PRA models. This handbook was generated to supplement these documents. It provides a compendium of good practices that a PRA analyst can use to generate the parameter distributions required for quantifying PRA models.

The increased use of risk assessment has also helped promote the idea that the collection and analysis of event data is an important activity in and of itself. In particular, the monitoring of equipment performance and evaluation of equipment trends can be used to enhance plant performance and reliability. The guidance provided in this handbook can support those efforts.

1.3 Scope

This handbook provides guidance on sources of information and methods for estimating parameter distributions. This includes determination of both plant-specific and generic estimates for initiating event frequencies, component failure rates and unavailabilities, and equipment non-recovery probabilities.

This handbook provides the basic information needed to generate estimates of the parameters listed above. It begins by describing the probability models and plant data used to evaluate each of the parameters. Possible sources for the plant data are identified and guidance on the collection, screening, and interpretation is provided. The statistical techniques (both Bayesian and classical methods) required to analyze the collected data and test the validity of statistical models are described. Examples are provided to help the PRA analyst utilize the different techniques.

This handbook also provides advanced techniques that address modeling of time trends. Methods for combining data from a number of similar, but not identical, sources are also provided. These are the empirical and hierarchical Bayesian approaches. Again, examples are provided to guide the analyst.

This handbook does not provide guidance on parameter estimation for all of the events included in a PRA. Specifically, common cause failure and human error probabilities are not addressed. In addition, guidance is not provided with regard to the use of expert elicitation. For these topics, the PRA analyst should consult other sources, such as the following references:

Common cause failures

- NUREG/CR-5497 (Marshall et al. 1998),
- NUREG/CR-6268 (Kvarfordt et al. 1998),
- NUREG/CR-5485 (Mosleh et al. 1998),
- NUREG/CR-4780 (Mosleh et al. 1988), and
- EPRI NP-3967 (Fleming, 1985).

Human errors

- NUREG/CR-1278 (Swain and Guttman, 1983),
- NUREG/CR-4772 (Swain, 1987),

- NUREG-1624 (NRC, 2000b), and
- EPRI TR-TR-100259 (Parry et al. 1992).

Expert Judgement

- NUREG/CR-6372 (Budnitz et al. 1997) and
- NUREG/CR-1563 (Kotra et al. 1996).

This list is not meant to be a comprehensive list of all of the methodologies available for performing these types of analyses.

1.4 Contents of the Handbook

This section provides a road map of the contents of the handbook and an overview discussion on how to use the handbook to perform the elements of a data analysis. The basics of probability and statistics described in Appendices A and B, respectively, are provided as reference material for the analyst. Appendix C provides statistical tables for selected distribution types that can be used in the data analysis.

1.4.1 Identification of Probability Models

The handbook provides guidance on the evaluation of five types of parameters that are included in a PRA:

- initiating events,
- failures to start or change state,
- failures to run or maintain state,
- durations, and
- unavailability from being out of service.

A description of each of these parameters along with examples, is provided in Chapter 2. Chapter 2 is fundamental reading for all users of this handbook.

The first step in a data analysis is to determine the appropriate probability models to represent the parameter. Chapter 2 provides a detailed description of the standard probability models for each event. This includes a discussion of the assumptions on the physical process inherent in the models and a description of the kind of data that can be observed. The type of data required to estimate the model parameter(s) are described and example data sets are examined in the light of the model assumptions. These examinations illustrate the kind of thinking necessary for the data analyst. Finally, a short discussion of related issues is presented for the analyst to consider.

1.4.2 Collection of Plant Specific Data

Once probability models have been defined for the basic events, plant-specific data should be evaluated for the purpose of quantifying estimates of the probability model parameters. Plant-specific data, if available in sufficient quantity and quality, is the most desirable basis for estimating parameter values. Chapter 5 discusses the process by which plant-specific data should be identified, collected, screened, and interpreted for applicability to the basic events defined in the systems analysis and to their probability models. To ensure that the collection and evaluation of plant-specific data is thorough, consistent, and accurate, the steps laid out in Chapter 5 should be followed for events defined in a PRA. The identification and evaluation of appropriate sources of plant-specific data for the basic events are discussed in Section 4.1.

The process for collecting and evaluating data for initiating events is discussed in Section 5.1. Guidance is provided for screening the data, for grouping the data into appropriate categories of initiating events, and for evaluating the denominator associated with the data.

The process for collecting and evaluating data for component failures is discussed in Section 5.2. It is critical that data be collected and processed accurately according to the definition of the component boundary. For example, it should be clearly noted whether or not a pump's control circuit is within or without the physical boundaries of the component for purposes of systems modeling. If failure of the control circuit has been modeled separately from hardware failures of the pump, then data involving failure of the pump should be carefully evaluated to ensure that actuation failures and other pump faults are not erroneously combined. This process could result in some iteration between the systems analysis task and the data collection task. It is possible that system models may be simplified or expanded based on insights derived during the data collection. Chapter 3 describes the difference between faults and failures, and discusses component boundary definitions and failure severity as it relates to data collection and analysis.

Other aspects of data collection for component failures discussed in Section 5.2 include classification and screening of the data, allocation of the data to appropriate component failure modes, and exposure evaluation (determining the denominator for parameter estimates).

The collection of data for recovery events is described in Section 5.3. Guidance is provided on where to find recovery-related data and on how to interpret such data.

1.4.3 Quantification of Probability Model Parameters

Once appropriate probability models have been selected for each basic event, estimates for the model parameters must be quantified. There are two basic approaches: 1) statistical estimation based on available data; and 2) utilization of generic parameter estimates based on previous studies. Both approaches can incorporate generic data. Several generic data sources currently available and used throughout the nuclear PRA industry are identified in Section 4.2.

1.4.3.1 Parameter Estimation from Plant-Specific Data

If the plant-specific data collection process yields data of sufficient quantity and quality for the development of parameter estimates, the statistical methods in Chapter 6 can be applied to the data to derive and validate parameter estimates for the basic events.

Chapter 6 discusses the statistical methods for estimating the parameters of the probability models defined in Chapter 2. Note that Appendix B discusses basic concepts of statistics that will help the user to understand the methods presented in Chapter 6.

For each type of event, two fundamental approaches are presented for parameter estimation: classical (frequentist) and Bayesian. An overview and comparison of these two approaches are presented in Section 6.1. The Bayesian approach is more commonly used in PRA applications, but classical methods have some use in PRA, as discussed in Section 6.1.

The probability models discussed in Chapter 2 for each type of event are applicable for most applications. However, erroneous results can occur in some cases if the assumptions of the model are not checked against the data. In some applications (e.g., if the impact of casual factors on component reliability is being examined) it is imperative that the probability model chosen for each basic event be validated given the available data. It may seem sensible to first confirm the appropriateness of the model and then estimate the parameters of the model. However, validation of a model is usually possible only after the model has been assumed and the corresponding parameters have been estimated. Thus, estimation methods are presented first in Chapter 6 for each type of probability model; then methods for validating the models against the available data are presented.

1.4.3.2 Parameter Estimation from Existing Data Bases

If actual data are unavailable or of insufficient quality or quantity then a generic data base will have to be used. Several generic data sources currently available and used throughout the nuclear PRA industry are identified in Section 4.2. Section 4.2.6 provides guidance on the selection of parameter estimates from existing generic data bases.

1.4.4 Advanced Methods

The last two chapters of the handbook describes some methods for analyzing trends in data and Bayesian approaches for combining data from a number of similar sources.

1.4.4.1 Analyzing Data for Trends and Aging

Data can be analyzed to assess the presence of time trends in probability model failure rates and probabilities (i.e., θ and p). Such trends might be in terms of calendar time or in terms of system age. Ordinarily, the analysis of data to model time trends involves complex mathematical techniques. However, the discussion of Chapter 7 presents various approaches that have been implemented in computer software. The discussion in Chapter 7 focuses on the interpretation of the computer output for application in PRA.

1.4.4.2 Parameter Estimation Using Data from Different Sources

Two Bayesian approaches for combining data from a number of similar, but not identical, sources are discussed in Chapter 8.

1.5 How to Use This Handbook

This handbook is intended for workers in probabilistic risk assessment (PRA), especially those who are concerned with estimating parameters used in PRA modeling. Broadly speaking, three groups of readers are anticipated: **data collectors**, who will be finding, interpreting, and recording the data used for the estimates; **parameter estimators**, who will be constructing the parameter estimates from the data and quantifying the uncertainties in the estimates; and (to a lesser extent) **PRA analysts**, who will be using the

estimated parameters. These three groups will find their primary interests in different portions of the handbook, as discussed below.

The major sections of the handbook can be grouped into several areas:

- Foundation: Chapters 1 and 2;
- Data Collection: Chapters 3, 4, and 5;
- Parameter Estimation: Chapters 6, 7, and 8; and
- Supporting Material: Appendices, References, Index.

These sections are shown in Figure 1.1, a schematic representation of the contents of the handbook.

PRA analysts will be most interested in the foundational material. Data collectors will need to read much of the foundational material, and then read the chapters on data collection. Parameter estimators will need to read the foundational chapters, but may then wish to skip directly to the relevant sections on parameter estimation. The supporting material can be read by anyone at any time.

The arrows in Figure 1.1 help the reader find the quickest way to the sections of interest. For example, the figure shows that Chapters 3-5 and Chapters 6-8 do not refer to each other or assume material from the other section, so it is possible to read from one section and not the other. The only strong dependencies are shown by the arrows: read Chapter 2 before starting Chapter 3 or 6, read Chapter 3 before starting Chapter 4 or 5, and so forth. In practice, data collectors, data analysts, and PRA analysts must work together, giving feedback to each other. The handbook, on the other hand, is formed of distinct segments, each of which can be read in isolation from the others.

The material for PRA analysts and data collectors is intended to be accessible by anyone with an engineering background and some experience in PRA. The material for data analysts, on the other hand, begins with elementary techniques but eventually covers advanced models and methods. These advanced topics will not be needed in most cases, but are included as reference material.

To aid the reader, Appendices A and B summarize the basics of probability and statistics, and Appendix C provides useful statistical tables. A glossary of terms is provided in Appendix D. Persons who have no previous experience with probability or statistics will need a more thorough introduction than is provided in these sections of the handbook.

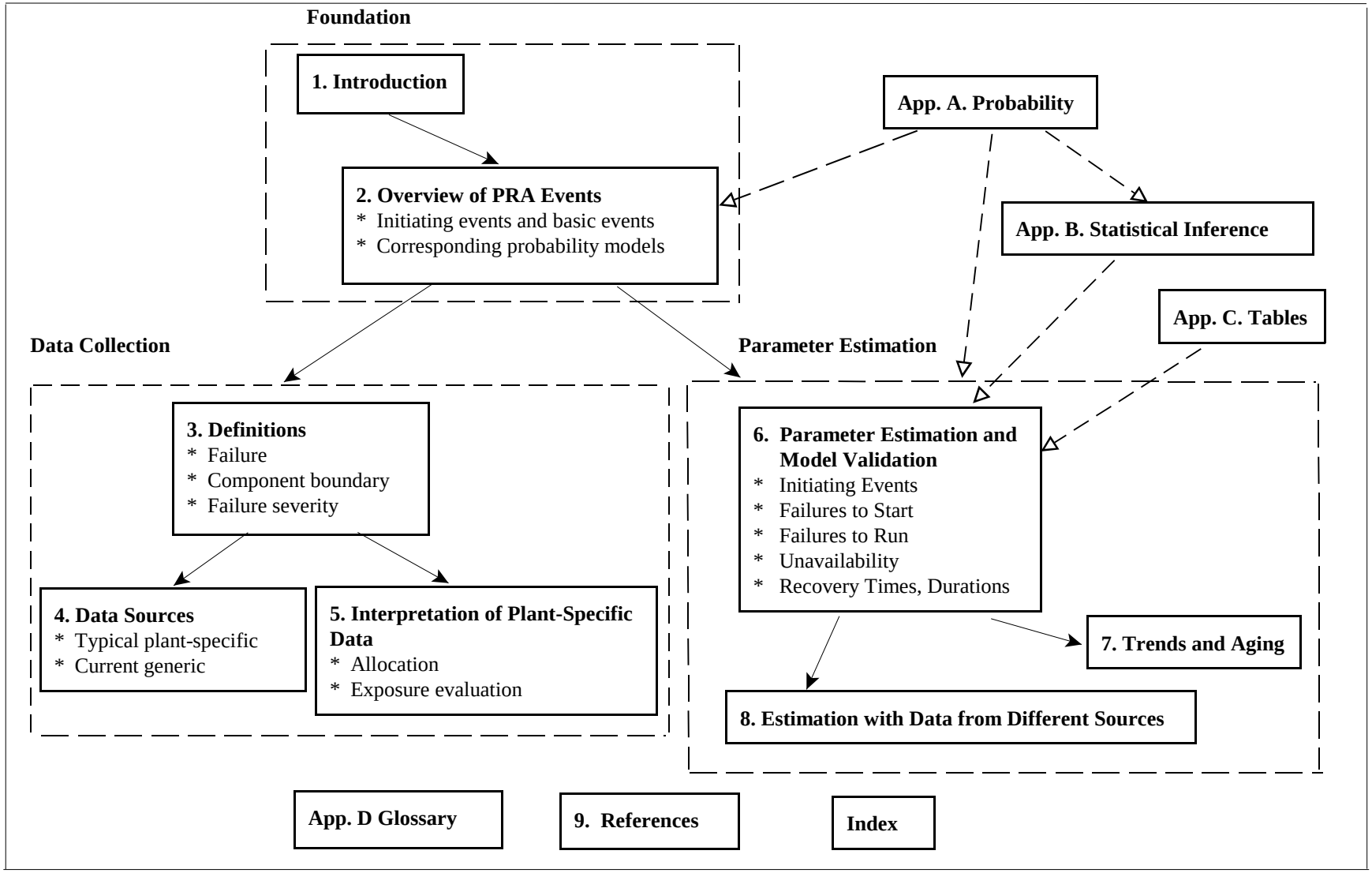


Figure 1.1 Relation of sections of this handbook. An arrow with a solid head indicates a strong dependence, with required information flowing in the direction of the arrow. An dashed arrow with a hollow head indicates that background or supplementary material flows in the direction of the arrow. The glossary, references, and index supply assistance for the entire handbook, so the arrows are not shown.

2. BASIC EVENT PROBABILITY MODELS

2.1 Overview

This chapter introduces the models used for basic events and for initiating events. This first section is an overview, and the remaining sections of the chapter give further detail.

Probabilistic risk assessment (PRA) considers various possible accident sequences. An accident sequence begins with an **initiating event** which challenges the safety of the plant. Typically, one or more standby safety systems are then demanded, and other, normally operating, systems must continue operating to ensure that no serious undesirable consequences occur. For the systems to fail to bring the situation under control, several components must either fail or be unavailable. The logic events in the PRA model that represent these failures or modes of unavailability are called **basic events**.

It is not possible to predict precisely when an initiating event or a component failure will occur, because the processes that lead to their occurrences are complex. Therefore, the initiating events and basic events are modeled as resulting from random processes.

The first step in the data analysis task is, therefore, to determine the appropriate probability model to represent the initiating event or basic event. (Probability is reviewed in Appendix A, and the probability models introduced here are presented more fully there.) These probability models typically have one or more parameters. The next major step is to estimate the values of these parameters. This estimation is based on the most applicable and available data. The process of choosing data sources, extracting the data in an appropriate form, and using it to estimate the parameters is the main subject of this handbook.

Basic events are customarily divided into unavailability (because the equipment is undergoing testing or maintenance), failure to start or change state, and failure to run (after successfully starting) or maintain state to the end of the required mission time. Unavailability and failure to run are each modeled in a single way. On the other hand, two different probability models have been used to represent a failure to start or to change state. The first method is to model the failures as having a constant probability of **failure on a demand**. The second method is to model the failures as occurring, in an unrevealed way, randomly in time. The failed condition is then discovered at the time of the demand. This is usually called the **standby failure-rate model**. Both models are discussed here.

The above events are the typical ones considered in a PRA. In addition, one must occasionally analyze durations, such as the time to restore offsite power or time to recover a failed component. Although such an analysis is not needed for a typical accident sequence, it is discussed in this handbook. Also, methods for analyzing durations can be used when estimating unavailability.

In summary, five topics are considered in the rest of this chapter:

- initiating events,
- failures to start or change state (modeled in two possible ways),
- failures to run or maintain state,
- durations, and
- unavailability from being out of service.

These topics are the subjects of Sections 2.2 through 2.6. Each section begins with examples of the data that might be analyzed. This is followed by a brief subsection presenting the assumptions of the usual model for the random process (the result of underlying physical mechanisms) and describing the kind of data that can be observed. The next subsection summarizes the data required to estimate the model parameter(s). The example data sets are then examined in the light of the model assumptions. These examinations illustrate the kind of thinking necessary for the data analyst. Finally, the section may conclude with a short discussion of related issues.

As a preview, Table 2.1 indicates the models, the parameters, and the data needed for each of the topics in the above five bullets. The top line of the table also indicates which section of Chapter 2 treats the topic.

The term **system** is used to denote the set of hardware for which data are collected; it may be an entire nuclear power plant (NPP), or a system in the traditional sense, such as the auxiliary feedwater (AFW) system, or a train, component, or even piece part. This reduces the need for phrases such as “system or component.”

The lengthiest part of each section below consists of the examination of examples to see whether the

Table 2.1 Kinds of models considered.

2.2 Initiating Events	2.3 Failures to Start or Change State (2 models)		2.4 Failures to Run or Maintain State	2.5 Durations	2.6 Unavailability
Typical Event					
Event occurs initiating an accident sequence	Standby system fails on demand		System in operation fails to run, or component changes state during mission	A condition persists for a random time period	System is unavailable, intentionally out of service, when demanded
Parameter(s) to Estimate					
δ, event frequency	For failure on demand: p, probability of failure on demand	For standby failure: δ, rate of occurrence of standby failures	δ, rate of failure to run	Parameters of assumed probability distribution of duration time	q, fraction of time when system will be out of service
Data Required to Estimate Parameters[a]					
Number of events, x, in total time, t	Number of failures, x, in total number of demands, n	Number of failures, x, in total standby time, t	Number of failures, x, in total running time, t	Depends on model, but typically the lengths of the observed durations	Onset times and durations of observed out-of-service events; OR observed fractions of time when system was out of service

[a.] The data here are the minimal requirements to estimate the parameter. More detailed data are needed to check the model assumptions.

assumptions of the probability model appear to be satisfied. Verifying model assumptions is an important part of good data analysis. Ways to investigate the appropriateness of assumptions are considered in Chapter 6, along with parameter estimation. The present chapter, however, only introduces the assumptions and illustrates their meanings through examples. If the assumptions are clearly not satisfied, some mention is given of ways to generalize the model, although such generalizations are not presented until Chapters 7 and 8 in this handbook.

Also, examples and extended discussion of examples are printed in Arial font, to distinguish them from the more general material.

2.2 Initiating Events

2.2.1 Examples

In the context of a nuclear-power-plant PRA, an initiating event is any event that perturbs the steady state operation of the plant, thereby initiating an abnormal event such as a transient or a loss-of-coolant accident within a plant. Initiating events begin sequences of events that challenge plant control and safety systems. Failure of these systems can lead to core damage and a release of radioactivity to the environment. However, the consideration of the potential plant response to initiating events is irrelevant when estimating their frequencies.

Here are several examples of data sets counting such initiating events.

Example 2.1 Unplanned reactor trips

A U.S. commercial nuclear power plant had 34 unplanned reactor trips in 1987 through 1995. It had its initial criticality on Jan. 3, 1987, and experienced a total of 64651 critical hours, or 7.38 critical years (Poloski et al. 1999a).

Example 2.2 Shutdown loss of offsite power

In U.S. commercial nuclear power plants in 1980-1996, there were 80 plant-centered loss-of-offsite-power (LOSP) events during shutdown. In that period, the plants experienced 455.5 reactor-shutdown years (Atwood et al. 1998).

Example 2.3 Through-wall pipe leaks

In world-wide experience of western-style pressurized water reactors (PWR)s (3362 calendar years of operation), a single through-wall leak event has been reported in large-diameter piping (Poloski et al. 1999a, Appendix J).

The final example of this section does not have initiating events in the usual sense. However, the model assumptions and the form of the data are exactly the same as for initiating events. Therefore, such data can be analyzed just like initiating-event data.

Example 2.4 Temperature sensor/transmitters

Eide et al. (1999a) report that temperature sensor/transmitters in the reactor protection system (RPS) of Westinghouse NPPs had 32 failures in 2264.1 component-years. These sensor/transmitters operate continuously, and when they fail they are repaired or replaced in a relatively short time. The number of failures is conservatively estimated from sometimes incomplete Nuclear Plant Reliability Data System (NPRDS) data, and the number of component years is based on an estimated number of components per loop.

These examples have several elements in common. First, they involve a **number of events** that occurred, and an **exposure time**, or **time at risk**, when the events could have occurred. The next subsection will present a simple probability model that gives rise to random events in time. In addition, in each of the above examples corrective action is taken after any event, so that the system then resumes operation (the system is **repairable**.) This means that the recorded operating history consists of a sequence of random event occurrences, which is summarized as a count of events in some fixed time. This type of data will direct us to a particular type of analysis, presented in Chapter 6.

The events may be the initiating events of an ordinary PRA (Example 2.1), initiating events of a shutdown PRA (Example 2.2), failures in a passive system (Example 2.3), which incidentally happen to be initiating events in a PRA. As mentioned above, Example 2.4 does not describe initiating events in the traditional PRA sense. However, the example may be analyzed in the same way as the first three examples, because the sensor/transmitter failures occur in a continuously running system and they initiate quick repair action. A PRA analyst would distinguish among the examples based on their safety consequences. The present discussion, however, adopts the viewpoint of probability modeling, in which the important fact is not the consequence of the events, but the way that they occur randomly in time. Reactor trip initiators are the prototypical examples of such events, but are not the only examples.

The exposure time is the length of time during which the events could possibly occur. In Example 2.1, the exposure time is reactor-critical-years, because a reactor trip can only occur when the reactor is at power. Because only one plant is considered, “critical years” can be used as shorthand for “reactor-critical-years.” In Example 2.2, the event of interest is LOSP during shutdown, so the exposure time must be the number of reactor-shutdown-years in the study period. In Example 2.3, reactor-calendar-years are used, primarily because more detailed worldwide data could not be easily obtained. The model therefore assumes that a crack in large-diameter piping could occur with equal probability during operation and during shutdown. The model also does not consider differences between plants, such as differences in the total length of large-diameter piping at a plant. In Example 2.4, the exposure time is the number of component-years, because the components operate constantly.

The possible examples are endless. The events could be unplanned demands for a safety system, forced outage events, or many other kinds of events that resemble initiating events.

The data given in the above examples are expressed in the crudest summary terms: a count of events in a total exposure time. This is sufficient for the simple model of this section. Section 2.5 will consider more sophisticated models using the exact event times.

The data could also be broken down into smaller pieces. For example, the initiating event data could be summarized for each calendar year, with an event count and an exposure time reported separately for each year from 1987 through 1995. This additional information allows one to look for trends or other patterns, as discussed in later chapters.

2.2.2 Probability Model

The assumptions concerning the physical process are given here, along with a description of the kind of data that can be observed.

It is standard to assume that the event count has a Poisson distribution. As listed in Section A.6.2, the usual assumptions (following Thompson 1981) for a Poisson process are:

1. The probability that an event will occur in any specified short exposure time period is approximately proportional to the length of the time period. In other words, there is a rate $\theta > 0$, such that for any interval with short exposure time Δt the probability of an occurrence in the interval is approximately $\theta \times \Delta t$.

2. Exactly simultaneous events do not occur.

3. Occurrences of events in disjoint exposure time periods are statistically independent.

In addition, it is worthwhile to spell out the kind of data that can be observed.

- A random number of events occur in some prespecified, fixed time period. As a minimum, the total number of events and the corresponding time period are observed.

Under the above assumptions, the number of occurrences X in some fixed exposure time t is a Poisson distributed random variable with mean $\mu = \theta t$,

$$\Pr(X = x) = e^{-\mu}\mu^{x} / x! \quad . \qquad (2.1)$$

The **probability distribution function** (p.d.f.) is sometimes used to abbreviate this: $f(x) = \Pr(X = x)$. (Throughout this handbook, upper case letters are used for random variables and lower case letters are used for particular numbers.)

The parameter θ is a **rate** or **frequency**. To make things more clear, the kind of event is often stated, that is, "initiating event rate" in Example 2.1, "through-wall-crack occurrence frequency" in Example 2.3, and so forth. Because the count of events during a fixed period is a unitless quantity, the mean number of occurrences μ is also unitless. However, the rate θ depends on the units for measuring time. In other words, the units of θ are per unit of time, such as 1/year or 1/reactor-critical-hour.

This model is called a **Poisson process**. It is extremely simple, because it is completely specified by the exposure time, t, and the one unknown parameter, θ. Assumption 1 implies that the rate θ does not change over time, neither with a monotonic trend, nor cyclically, nor in any other way. Assumption 2 says that exactly simultaneous events do not occur. The only way that they could occur (other than by incredible coincidence) is if some synchronizing mechanism exists – a common cause. Therefore, the operational interpretation of Assumption 2 is that common-cause events do not occur. Assumption 3 says that the past history does not affect the present. In particular, occurrence of an event yesterday does not make the probability of another event tomorrow either more or less likely. This says that the events do not tend to occur in clusters, but nor do they tend to be systematically spaced and evenly separated.

As stated above, a common cause that synchronizes events violates Assumption 2. However, some common-cause mechanisms do not exactly synchronize the events. Instead, the second event may occur very soon after the first, as a slightly delayed result of the common cause. In this case, Assumption 3 is violated, because the occurrence of one event increases the probability of a second event soon after. One way or the other, however, common-cause events violate the assumptions of a Poisson process, by violating either Assumption 2 or Assumption 3.

2.2.3 Data Needed to Validate the Model and Estimate θ

Suppose that the Poisson model holds. Then any reasonable estimator of θ needs only two pieces of information: the total exposure time, t, in the data period, and the number of events, x, that occurred then.

However, more information is needed to investigate whether the Poisson model is valid. For example, the data might cover a number of years or a number of plants, and δ might not be constant over time or the same at all plants. These possibilities are not allowed by the listed model assumptions. To study whether they occur, the times and locations of the initiating events should be recorded, or at least the data should be partitioned into subsets, for example corresponding to plants or years. Then the event count and exposure time, x_i and t_i, should be given for each subset.

2.2.4 Case Studies: Validity of Model Assumptions in Examples

Let us examine the reasonableness of the Poisson model assumptions for Examples 2.1 through 2.4. Chapter 6 will address this issue by performing data analysis. Here we will merely cite the results of published studies and use critical thinking.

Example 2.1 Initiating Events

An initiating event is an event with the reactor critical, causing an unplanned reactor trip. Assume that any time interval starts on some date at some time and ends on some date at some time, and that the length of the interval, τ, is the number of critical years contained between the start and stop of the time interval. For example, if the time period is two 24-hour days and the reactor was critical for half of that time, then τ = 1/365 critical years.

Assumption 1 is violated in two ways. First, in the industry as a whole, and presumably in individual plants, the probability of an initiating event in an interval of length τ (such as one critical day) has not been constant. Instead, the probability dropped substantially from 1987 to 1995. Equivalently, the event rate, δ, dropped from 1987 to 1995. This violation can be eliminated by considering only a short time period for the study, such as one calendar year instead of nine years. If, however, the whole nine-year period is of interest, a more complicated model must be used, such as one of the trend models described in Chapter 7.

A second violation of Assumption 1 arises because this particular plant was new at the start of the study period, with initial criticality on January 3, 1987, and commercial start on May 2, 1987. Many new plants seem to experience a learning period for initiating events, and this plant had 15 of its 34 initiating events during the first six months of 1987. After that initial period with a high event rate, the event rate dropped sharply. This violation of Assumption 1 can be resolved by eliminating data before the plant reached a certain age. That is, not counting either the operating time or the initiating events from the plant until it has reached a certain age — excluding that portion of the plant's history from the universe being studied.

Assumption 2 says that exactly simultaneous initiating events do not occur. This is reasonable for events at a single plant.

Assumption 3 says that the probability of an initiating event in one time period does not depend on the presence or absence of an initiating event in any earlier time period. This assumption may be challenged if the plant personnel learn from the first event, thus reducing the probability of a second event. This kind of dependence of one event on another is not allowed by Assumption 3. Suppose, however, that the learning is modeled as a general kind of learning, so that the event rate decreases over time but not as a clear result of any particular event(s). This may justify using a Poisson model with a trend in the event rate, as considered in detail in Chapter 7.

There is a length of time when the reactor is down after a reactor trip when an initiating event cannot possibly occur. This does not violate Assumption 3 because during that time the plant has dropped out of the study. Its shutdown hours are not counted in the exposure time. Only when the reactor comes up again does it begin contributing hours of exposure time and possible initiating events.

Example 2.2 Shutdown LOSP

Just as with the previous example, consider the three assumptions of the Poisson model. In this case, because data come from the entire industry, δ is interpreted as the average rate for the entire industry.

First consider Assumption 1. The report that analyzed this data (Atwood et al. 1998) found no evidence of a trend in the time period 1980 through 1996. It did find evidence of differences between plants, however. These differences can affect the industry average, because plants enter the study when they start up and leave the study when they are decommissioned. When a plant with an especially high or low event rate enters or leaves the study, this will affect the industry average. However, the event rate at the worst plant differed from the industry average by only a factor of about 3.4, and the best plant differed from the average by less than that. Many plants (116) were considered. Therefore, the effect of a single plant's startup or decommissioning should be small. Therefore, it appears that the overall industry event rate was approximately constant, as required by Assumption 1.

Assumption 2 rules out exactly simultaneous events. In this example, however, events at sister units at a single site are somewhat dependent, because a common cause can result in LOSP events that are simultaneous or nearly simultaneous at both units.

Of the 80 events in the data, two pairs of events occurred together at sister units, each pair from a common cause. Thus, simultaneous events do occur, but they are not frequent. This departure from Assumption 2 is probably not large enough to be serious.

Assumption 3 requires statistical independence of the number of events in disjoint time intervals. As with Example 2.1, there may be some learning, although the lack of trend indicates that any learning is minimal.

In summary, the assumptions for the Poisson model seem to be approximately satisfied.

Example 2.3 Through-Wall Leaks

This differs from the other examples in that the number of events is very small. Any departures from the Poisson assumptions cannot be seen in the data, because so few events have occurred. With no theoretical reason to postulate a trend or other nonconstancy, or a high rate of multiple events, or dependence between events, we accept the Poisson assumptions. The assumptions may not be perfectly true, and a different model may be more accurate, but the Poisson model is simple, and good enough for analyzing such a sparse data set.

Example 2.4 Temperature Sensor/Transmitters

A report by Eide et al. (1999a) divides the total study time for instrumentation failures into two halves, and finds a difference between λ in 1984-1989 and λ in 1990-1995. The example here is for 1990-1995 only. Within this time period the report does not see strong evidence of a trend. That is, a small trend may be present, but the time period is too short, and the failures too few, for any trend to be clear. Further, because the components are regularly maintained, it is reasonable to assume that the failure rate, λ, is roughly constant, as required by Assumption 1.

Assumption 2 requires that common-cause failures be negligible. However, the report states that 14 of the 32 component failures occurred during four common-cause events. Thus, Assumption 2 is seriously violated.

Finally, Assumption 3 requires independence of the number of events in disjoint time intervals. The report does not address this issue, but independence appears plausible.

In summary, the example violates Assumption 2, but probably satisfies the other two assumptions. One way to deal with the violation of Assumption 2 would be to model the independent failures and the common-cause failures separately, although Eide et al. do not do this.

2.2.5 Discussion

2.2.5.1 More General Models

The model considered thus far is a **homogeneous Poisson process (HPP)**, which has a constant event occurrence rate, λ. The number of events in time t is a Poisson random variable with parameter $\mu = \lambda t$. A generalization is a **nonhomogeneous Poisson process (NHPP)**, in which λ is a function of t. Such a model is useful for analyzing trends. Chapter 6 includes ways to test the assumptions of a homogeneous Poisson process, and Chapter 7 includes ways to analyze data where a trend is present.

When data come from the industry, one may consider the differences between plants. Ways to model such differences are discussed in Chapter 8 of this handbook. The present chapter's interest is restricted to λ when no such variation is present. Of course, if the data come from only one plant, λ refers to that plant and the issue of differences typically does not arise.

Any mathematical model, such as the model for a homogeneous Poisson process given here, is an imperfect approximation of the true process that generated the data. Data are used to validate or refute the adequacy of the model. The data set may be sparse — in the present context, this means that the data set contains few events. In this case, two consequences typically result: (a) it is difficult or impossible to see evidence of departures from the model, and (b) the data set contains too little information to allow realistic estimation of the parameters of a more complicated model. If, instead, the data set has many events, departures from the model become visible, and typically a more complicated model is appropriate. These statements have been illustrated by the small and large data sets given as examples.

2.2.5.2 Non-randomness of t

In the model considered here, the exposure time is treated as fixed, and the number of events is treated as random. This is a common type of data found in PRA work. Sometimes, however, a fixed number of events is specified by the data collector, and the corresponding total time is random, as in the following two examples.

One example occurs when equipment is tested until it fails. That is, a predetermined number of items are tested, say x items. Each item is run until it fails, and the total running time of the items is random. The second example occurs in a PRA context if the analyst thinks that the event frequency has changed over time and that only the recent history fully represents current conditions. The analyst may then decide to consider only the most recent events. If there are four recent events, x is fixed at 4, and the corresponding time, measured backwards from the present to the 4th event in the past, is random.

These are examples of **duration** data with exponentially distributed durations, discussed in Section 2.5. The probability model is the Poisson process presented above, but the data collection, and resulting data analysis, are different. Because the time t until the xth event can be called a waiting time, these models are also sometimes called **waiting time models**.

2.3 Failure to Change State

This section considers two probability models, in Subsections 2.3.2 and 2.3.3. First, however, example data sets are given.

2.3.1 Examples

Here are four examples of failure to change state, three with failure to start and one with failure to close.

Example 2.5 HPCI failures to start

At 23 BWRs in the 1987-1993 time period, the high pressure coolant injection (HPCI) system had 59 unplanned attempts to start. The system failed to start on 5 of these demands (Grant et al. 1999a). The failures were typically erratic starts, which the operator stabilized manually. These demands occurred during 113.94 reactor-critical-years.

Example 2.6 EDG failures to start

Emergency diesel generators (EDGs) are sometimes demanded because of unplanned loss of power to a safety bus, and they are also tested periodically, with one set of tests during each operating cycle and another set of tests monthly. In addition, a return-to-service test is normally performed after maintenance of an EDG. At one plant over an 18-month time period, the number of such demands is counted, and the number of failures to start is counted.

Example 2.7 Steam binding in AFW

Between demands, steam binding can develop in the AFW system, so that one or more pumps cannot function when demanded. This is mentioned by Wheeler et al. (1989), and by Nickolaus et al. (1992).

Example 2.8 Failures of isolation valves

Nickolaus et al. (1992) review the causes of about 45 failures of air-operated and motor-operated isolation valves. Some of the principal causes are corrosion, instrument drift, and moisture in instrument and control circuits. Other causes include contamination and corrosion products in the instrument air system, and debris in the system. These are all conditions that can develop while the valves are not being used.

2.3.2 Failure on Demand

All these examples involve a number of demands and a number of failures, where the terms “demand” and “failure” can be defined according to the purposes of the study. Non-PRA contexts provide many other examples of failures on demand. A simple example in elementary probability or statistics courses is tossing a coin n times, and counting the number of heads. Count either a head or a tail as a “failure.” Just as in the PRA examples, this example has a number of demands, with a random number of the demands resulting in failures.

2.3.2.1 Probability Model

The standard model for such data assumes that the number of failures has a binomial distribution. The assumptions are listed in Appendix A.6.1. These assumptions can be restated as two assumptions about the physical process and one about the observable data:

1. On each demand, the outcome is a failure with some probability p, and a success with probability 1 ! p. This probability p is the same for all demands.

2. Occurrences of failures for different demands are statistically independent; that is, the probability of a failure on one demand is not affected by what happens on other demands.

The following kind of data can be observed:

- A random number of failures occur during some fixed, prespecified number of demands. As a minimum, the total number of failures and number of demands are observed.

Under these assumptions, the random number of failures, X, in some fixed number of demands, n, has a binomial (n, p) distribution.

$$\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, \ldots, n \tag{2.2}$$

where

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} .$$

This distribution has two parameters, n and p, of which only the second is unknown. (Although n may not always be known exactly, it is treated as known in this handbook. Lack of perfect knowledge of n, and other uncertainties in the data, are discussed briefly in Section 6.1.3.2)

2.3.2.2 Data Needed to Validate the Model and Estimate *p*

Suppose that the binomial model is appropriate. Then any reasonable estimator of p needs only two pieces of information: the number of demands, n, in the data period, and the number of failures, x, that then occurred.

However, more information is needed to investigate whether the binomial model is valid. For example, Assumption 1 assumes that p is the same on all demands. If the data cover a number of years or a number of systems or plants, p might not be constant over time or the same at all systems or plants. To study whether this is true, the times and locations of the demands and failures should be recorded, or at least the data should be partitioned into subsets, for example corresponding to systems, plants, or years. Then the failure and demand counts, x_i and n_i, should be given for each subset.

2.3.2.3 Case Studies: Validity of Model Assumptions in Examples

Let us examine Examples 2.5 through 2.8 to see if the assumptions appear to be valid.

Example 2.5 HPCI Failures to Start

Assumption 1 says that the probability of failure on demand is the same for every demand. If data are collected over a long time period, this assumption requires that the failure probability does not change. Likewise, if the data are collected from various plants, the assumption is that p is the same at all plants.

In the HPCI example, the five failures do not reveal any clear trend in time. However, one Licensee Event Report (LER) mentions that a better-designed switch had already been ordered before the HPCI failure. This gives some evidence of a gradual improvement in the HPCI system, which might be visible with more data.

As for differences between plants, it happens that three of the five failures occurred at a single plant. Therefore, it might be wise to analyze that one plant (three failures in nine demands) separately from the rest of the industry (two failures in 50 demands). In fact, Grant et al. (1995) did not analyze the data that way, because they considered two types of failure to start, and they also considered additional data from full system tests performed once per operating cycle. However, the high failure probability for the one plant was recognized in the published analysis.

Assumption 2 says that the outcome of one demand does not influence the outcomes of later demands. Presumably, events at one plant have little effect on events at a different plant. However, the experience of one failure might cause a change in procedures or design that reduces the failure probability on later demands at the same plant. One of the five LERs mentions a permanent corrective action as a result of the HPCI failure, a change of piping to allow faster throttling. This shows some evidence of dependence of later outcomes on an earlier outcome at that plant.

Example 2.6 EDG Failures to Start

Assumption 1 says that every demand has the same probability, p, of failure. This is certainly not true for return-to-service tests, because such tests are guaranteed to result in success. If the EDG does not start on the test, maintenance is resumed and the test is regarded as a part of the maintenance, not as a return-to-service test. Therefore, any return-to-service tests should not be used with the rest of the data.

As for the other demands, one must decide whether the unplanned demands, operating-cycle tests, and monthly tests are similar enough to have the same value of p. Can plant personnel warm up or otherwise prime the diesel before the test? Can an

operator stop the test if the EDG is clearly having trouble, and then not consider the event as a test? If so, the different types of demands do not have the same p, and they should not be analyzed as one data set. For PRA purposes, one is normally most interested in the failure probability on an actual unplanned demand. To estimate this, one should use only data from unplanned demands and from tests that closely mimic unplanned demands.

If the EDGs in the data set differ in some way, such as having different manufacturers, this may also lead to different values of p on different demands. Analyzing the data while ignoring differences between the individual EDGs will allow us to estimate the average p, corresponding to failure to start for a random EDG. However, this average p is not the same as the p for a particular EDG.

Assumption 2 says that the outcome on one demand does not affect the probability of failure on a different demand. When the plant is very new there may be some learning from individual failures, but when the plant is mature, failure or success on one demand should not change the chances of failure or success on later demands. The only way for such dependence to arise is if the first failure results from a common cause. If the plant is mature and common-cause failures are rare, then Assumption 2 is approximately satisfied.

Example 2.7 Steam binding in AFW

Assumption 1 says that every demand corresponds to the same probability of failure. If the steam comes from backflow through a check valve, it will build up, and become more of a problem when the AFW system has been unattended longer. Technically, this is a violation of Assumption 1. However, ignoring the differences between demands results in estimating p for an average demand, and this may be adequate for many purposes.

Assumption 2 says that the AFW pumps fail independently of each other. However, steam-binding of the AFW system was a recognized common-cause mechanism in the 1970s and 1980s. This means that Assumption 2 may be plausible if interest is in the performance of a single AFW pump, but not if interest is in an interconnected set of pumps.

Section D-1 of Poloski et al. (1998) says that steam binding has not been seen in 1987-1995 AFW experience. Therefore, Example 2.7 is probably no longer relevant, although it received great attention at one time.

Example 2.8 Failures of isolation valves

The causes of valve failures postulated in this example are degradations, so the probability of failure increases over time, violating Assumption 1. If failures from such causes are rare, then the increase in failure probability may not be a problem. In general, ignoring the differences results in estimating an average p, averaged over components that have been allowed to degrade for different amounts of time. This may be acceptable.

As in Example 2.7, some of the mechanisms for valve failure are common causes, violating the independence required by Assumption 2. The seriousness of the violation depends on how many multiple failures occur.

2.3.2.4 Discussion

2.3.2.4.1 More General Models

The model considered above has a constant failure probability, p. A generalization would let p be a function of time. Such a model is useful for analyzing trends. Chapter 6 includes ways to test the assumptions of the model assumed above, and Chapter 7 includes ways to analyze data where a trend is present.

When data come from the industry, one might consider the differences between plants, just as for initiating events. Ways to model such differences are discussed in Chapter 8. The present section's interest is restricted to p for the industry as a whole, the average of all the plants. Of course, if the data come from only one plant, p refers to that plant and the issue of differences typically does not arise.

Any mathematical model is an imperfect approximation of the true process that generated the data. When the data set is sparse (few demands, or few or no failures, or few or no successes), (a) it is difficult or impossible to see evidence of departures from the model, and (b) the data set is too small to allow realistic estimation of the parameters of a more complicated model. When the data set has many events, departures from the model become visible, and a more complicated model may be appropriate.

2.3.2.4.2 Non-randomness of n

One could argue that the numbers of demands in the examples are not really fixed in advance. That is, no one decided in advance to look at the outcomes of 59 unplanned HPCI demands. Instead, Grant et al. decided to look at seven years of data from 23 plants,

and they observed that 59 demands had taken place. The response to this argument is that we are actually conditioning on the number of demands, that is, dealing with conditional probabilities assuming that 59 demands take place. Conditioning on the number of demands enables us to focus on the quantity of interest, p. Treating both the number of failures and the number of demands as random is needlessly complicated, and yields essentially the same conclusions about p as do the simpler methods in this handbook.

In the model considered here, the number of demands is treated as fixed, and the number of failures is treated as random. Sometimes, however, the number of failures is specified in advance and the corresponding number of demands is random. For example, the analyst may believe that p has been changing, and that only the most recent history is relevant. In this case, the analyst might decide to consider only the most recent failures and to treat the corresponding number of demands as random. For example, if only the four most recent failures are included, one would count backwards from the present until $x = 4$ failures were seen in the plant records, and record the corresponding number of demands, n, regarded as an observation of a random variable. This is a **waiting time model**, with n equal to the waiting time until the 4th failure. Bayesian analysis of such data is discussed briefly in Section 6.3.2.6.

2.3.3 Standby Failure

As stated in the introduction to this chapter, failure to change state can be modeled in two ways. One way was given in Section 2.3.2. The second way is given here, in which the system (typically a component) is assumed to transform to the failed state while the system is in standby. This transition occurs at a random time with a constant transition rate. The latent failed condition ensures that the system will fail when it is next demanded, but the condition is not discovered until the next inspection, test, or actual demand.

2.3.3.1 Probability Model

The underlying assumption is that the transition to the failed condition occurs randomly in time. Two settings must be distinguished:

1. the *data*, the operational experiences in the past that allow us to estimate θ, and
2. the *application* to PRA, in which the estimate of θ is used to estimate the probability that a component will fail when demanded.

These two settings are discussed in the next two subsections.

2.3.3.1.1 Probability Model for the Data

It is customary to consider only the simplest model.

1. Assuming that the system is operable at time t, the probability that the system will fail during a short time period from t to $t + \Delta t$ is approximately proportional to the length of the exposure period, Δt. The probability does not depend on the starting time of the period, t, or on anything else.

2. Failures of distinct systems, or of one system during distinct standby periods, are independent of each other.

The kind of observable data is spelled out here. It is obvious, but is written down here for later comparison with the data for similar models.

- At times unrelated to the state of the system, the condition of each system (failed or not) can be observed. As a minimum, the total number of failures and the corresponding total standby time are observed.

The times mentioned here can be scheduled tests or unplanned demands.

Assumption 1 is essentially the same as for a Poisson process in Section 2.2.2. It implies that there is a proportionality constant, θ, satisfying

$$\theta \Delta t \approx \Pr(t < T \le t + \Delta t \mid T > t),$$

where T is the random time when the system becomes failed. Then the probability that the system is failed when observed at time t is

$$\Pr(\text{system is in failed state at time } t) = 1 - e^{-\theta t}. \quad (2.3)$$

This follows from Equation 2.6, given in Section 2.5 for the exponential distribution. The parameter θ is called the **standby failure rate**. It is so named because the failed condition develops while the system is in standby, waiting for the next demand.

2.3.3.1.2 Application of the Model to PRA

The model is used to evaluate the probability of failure on an unplanned demand. For this, one assumes that there are periodic tests and the unplanned demand occurs at a random time within the testing cycle. Then the probability of failure on demand is approximated by

$$p \approx \lambda t_{test}/2 \quad , \tag{2.4}$$

where λ is the standby failure rate and t_{test} is the time interval between tests.

A more accurate expression is the average of terms from Equation 2.3, averaging over all the possible demand times in the test interval:

$$p = \frac{1}{t_{test}} \int_0^{t_{test}} \left(1 - e^{-\lambda s}\right) ds$$
$$= 1 - \left(1 - e^{-\lambda t_{test}}\right) / (\lambda t_{test})$$

This equation is approximated by Equation 2.4, as can be verified by use of the second-order Taylor expansion:

$$\exp(-\lambda t) \approx 1 + (-\lambda t) + (-\lambda t)^2/2! \approx .$$

When more than one system is considered, the formulas become more complicated. For example, suppose that two systems (such as two pumps) are tested periodically and at essentially the same time. Suppose that we are interested in the event that both fail on an unplanned demand. This is:

$$\Pr(\text{both fail})$$
$$= \frac{1}{t_{test}} \int_0^{t_{test}} (1 - e^{-\lambda s})^2 ds \tag{2.5}$$
$$\approx (\lambda t_{test})^2 / 3$$

When more systems are involved, or when testing is staggered, the same ideas can be applied.

2.3.3.2 Data Needed to Validate the Model and Estimate λ

Suppose that the standby failure rate model holds. If the standby times are all similar, then an estimator of λ needs only two pieces of information: the number of failures, x, in the data period, and the corresponding total standby time, t. If, instead, the standby times vary substantially, then the total standby times should be recorded separately for the failures and the successes, as explained in Section 6.4.

To validate the model, the data could be partitioned. As with initiating events, if the data come from various years or plants, the data could be partitioned by year and/or by plant, and the above information should be given for each subset.

2.3.3.3 Case Studies: Validity of Model Assumptions in Examples

Let us now examine the applicability of the model assumptions in the examples given above. Much of the discussion in Section 2.3.2.3 applies here as well. In particular, when Section 2.3.2.3 sees a violation of an assumption and suggests a remedy, an analogous violation is probably present here, with an analogous remedy.

Example 2.5 HPCI Failures to Start

Assumption 1 says that the probability of becoming failed in a short time period is proportional to the length of the time period, and on nothing else. As discussed in Section 2.3.2.3, there is no clear evidence of a trend in time. It may be, however, that the probability of failure is higher at one plant than at the other plants. If true, this would violate Assumption 1, and suggests that the outlying plant be analyzed separately from the others.

Assumption 2 says that failures in distinct time periods and locations are independent of each other. As discussed in Section 2.3.2.3, there may be a very small amount of learning, causing fewer failures later in the history.

Example 2.6 EDG Failures to Start

Assumption 1 says that the probability of becoming failed in a short time period is proportional to the length of the time period, and on nothing else. Section 2.3.2.3 discusses different types of tests of EDGs. That discussion is applicable here as well. If an EDG fails on one type of test more readily than on another type of test, Assumption 1 is violated. Another interpretation of this situation is that the bulleted assumption on the data is false: it is not true that a failed condition is always discovered on a test. Some tests discover only major failed conditions while other, more demanding tests discover less obvious failed conditions. Just as mentioned in Section 2.3.2.3, if the primary interest is the probability of failure on an unplanned demand then one should use only data from unplanned demands and from tests that closely mimic unplanned demands.

Assumption 2 says that failures in distinct time periods and locations are independent of each other. As discussed in Section 2.3.2.3, this is probably true if the plant is mature and if common-cause failures are rare.

Example 2.7 Steam Binding in AFW

Assumption 1 says that the failed-condition event is as likely to hit the system in one time interval as in another of the same length. As discussed in Section 2.3.2.3, steam binding can result from a gradual buildup, and become more of a problem when the AFW system has been unattended longer. In this case, Assumption 1 is violated. Ignoring this fact is equivalent to treating the average of AFW conditions.

As discussed in Section 2.3.2.3, steam binding is a common-cause mechanism. Therefore Assumption 2, independence of distinct AFW pumps, is violated.

Example 2.8 Failures of Isolation Valves

Just as discussed in Section 2.3.2.3, the causes listed for Example 2.3 are degradations, violating Assumption 1. However, it may be acceptable to ignore the changes over time, and estimation of an average parameter δ. Also, as discussed in Section 2.3.2.3, some of the mechanisms for valve failure are common causes, violating the independence required by Assumption 2. The seriousness of the violation depends on how many multiple failures occur.

2.3.4 Comparison of the Two Models for Failure to Change State

Two models have been presented for failure to change state, the failure-on-demand model and the standby-failure model. Several aspects of the models are compared here.

2.3.4.1 Ease of Estimation

One great appeal of the standby-failure model is that the analyst does not need knowledge of the number of demands. Standby time is normally much easier to obtain than a count of demands.

2.3.4.2 Use in PRA Cut Sets

The two models differ in their application to cut sets in a PRA model. Consider failure of two redundant components, each having the same probability of failure. When the failure-on-demand model is used, we have

$$\Pr(\text{both fail}) = p^2 = [\Pr(\text{one fails})]^2.$$

On the other hand, when the standby-failure model is used and the two components are tested periodically at the same time, with time t between tests, Equations 2.4 and 2.5 show that

$$\Pr(\text{one fails}) \approx \delta t_{test}/2$$

$$\Pr(\text{both fail}) \approx (\delta t_{test})^2/3$$

so that

$$\Pr(\text{both fail}) \ll [\Pr(\text{one fails})]^2.$$

This fact is often ignored.

2.3.4.3 Estimates Obtained

The two models can produce different estimates of basic event probabilities. For example, suppose that an EDG is tested monthly by starting it. In 100 monthly tests, 2 failures have been seen. A simple estimate of p, the probability of failure on demand, is 2/100 = 0.02. A simple estimate of δ, the standby failure rate, is 0.02/month. Now suppose that a basic event in a PRA is that the EDG fails to start, when demanded at a random time. Based on the estimate of p, the estimated probability of the basic event is

$$\Pr(\text{EDG fails to start}) = p \approx 0.02 .$$

Based on the estimate of δ and Equation 2.4, the estimated probability of the basic event is

$$\Pr(\text{EDG fails to start}) \approx \delta t/2$$
$$\approx (0.02/\text{month})\times(1\ \text{month})/2 = 0.01 .$$

The two models give estimates that differ by a factor of two, with the failure-on-demand model being more pessimistic than the standby-failure model. The reason is simple: All, or virtually all, of the failures and demands in the data occur at the end of test intervals. However, unplanned demands might occur at any time between tests. The standby-failure model says that demands soon after a successful test have smaller probability of failure. The failure-on-demand model says that all demands have the same probability of failure.

The differences can be more extreme. For example, suppose that two EDGs are tested monthly, and tested at essentially the same time rather than in a staggered way. According to the failure-on-demand model, the probability that both EDGs fail to start is p^2, which is estimated by $(0.02)^2$. On the other hand, according to the standby-failure model, Equation (2.5) shows that the same probability is approximately $(\delta t_{test})^2/3$, which is estimated by $(0.02)^2/3$. The two models give estimates that differ by a factor of three. More extreme examples can be constructed.

It might be mentioned that these numerical differences between estimates disappear if only unplanned demands are used in the data. However, unplanned demands are rare, and so most analysts prefer to use test data if possible.

2.3.4.4 A Model Involving Both Terms

The model described next postulates two reasons for the observed randomness of failures.

One reason for the randomness of failures is that demands are not all equally stressful. When a demand occurs that is unusually harsh, the system will fail. From the viewpoint of an outside observer, it appears that failures just occur randomly with some probability p, but the underlying cause is the variability in the severity of the demands.

The other reason for randomness of the failures is that the unattended system degrades, and becomes inoperable at unpredictable times. This is simplified in the standby-failure model by supposing that the system changes suddenly from perfectly operable to completely failed, with these transitions occurring at random times. This leads to the standby-failure model, with failure-transition rate δ, and with probability of failure δt at time t after the last system restoration.
If just one of the two mechanisms described above is considered, we are led to either the failure-on-demand model or the standby-failure model. It is possible, however, to construct a model that involves both terms, corresponding to the two kinds of variation. In this two-parameter model, the probability of failure is $p + \delta t$ at time t after the last system restoration. (For example, see Section 5.2.10 of Samanta et al. 1994.)

Lofgren and Thaggard (1992) state "it is virtually impossible to directly determine from work maintenance record descriptions whether the component has failed from standby or demand stress causes." However, they look for patterns in data from EDGs and motor-operated valves (MOVs) at a small number of plants that use different test intervals. Their data suggest that the standby-failure-rate model is most appropriate for MOV failures, and the two-parameter model is best for EDGs.

In a similar spirit, the T-Book (TUD Office and Pörn Consulting, 2000) uses the two-parameter model for many components. The T-Book does not attempt to identify which mechanism applies to which failures, but instead estimates the two parameters from overall patterns in the data. Some of the resulting estimates have large uncertainties; for example, at a typical plant the estimate of p for EDG failure to start has an error factor of about 13. For components that cannot be analyzed in this way, the T-Book uses the standby-failure model. For details, see Pörn (1990).

2.3.4.5 Choosing a Model

No consensus exists among PRA workers as to which model is most advantageous. In particular, the typical mechanisms of failure are not understood well enough to justify a theoretical basis for a model. Most current work uses one of the two simple models given here: failure on demand or standby failure. Therefore, this handbook presents only these two models. The user may choose between them.

2.4 Failure to Run during Mission

Aspects of this type of failure closely resemble the initiating events of Section 2.2. One important difference is in the kind of data normally present. The difference is summarized here.

Example 2.4 of Section 2.2 is an example of continuously running components (temperature sensor/transmitters) that occasionally failed to run. When a component failed, it was repaired or replaced in a relatively short time, and resumed operation. That is, the component was **repairable**. The present section considers components or systems that do not run continuously. Instead, they are occasionally demanded to start, and then to run for some mission time. If they fail during the mission, they are **nonrepairable**, that is, they cannot be repaired or replaced quickly. Two points deserve clarification:

- Some failures may be recoverable. They would not be modeled as failures in the sense of causing mission failure. Unrecoverable failures cause mission failure, however.

- Given enough time, almost any system can be repaired. During a mission, however, time is not available. Because the component or system cannot be repaired *within the time constraints*, it is called "nonrepairable."

As stated earlier, the word **system** is used in this handbook for any piece of hardware for which data are taken. In particular, components and trains are kinds of systems.

2.4.1 Examples

Here are two examples of failures to run during missions.

Example 2.9 EDG failures to run

Grant et al. (1999b) report that in 844 demands of 30 minutes or more for EDGs to run, there were approximately 11 unrecovered failures to run in the first 30 minutes. The count is approximate because a few failure times were not given and had to be inferred.

Example 2.10 AFW turbine train failures to run

Poloski et al. (1998) report that in 583 unplanned demands of AFW system turbine trains, the train failed to run 2 times, and the total running time was 371 train-hours. The information is taken from LERs, only 17% of which report running times for the train. The total running time of 371 hours is an extrapolation from the LERs with reported run times.

These examples are typical, in that hardly any of the demands to run resulted in a failure. Therefore, for most demands the time when failure would eventually have occurred is unknown.

2.4.2 Probability Model

In principle, the times to failure are **durations**. Section 2.5 deals with duration data, in the context of recovery times. That section mentions various possible distributions of time to failure, of which the simplest is the exponential distribution.

Data for this section differ from data of Section 2.5, however, because nearly all of the observed times in this section are truncated before failure. This is illustrated by the above examples. Therefore, the full distribution of the time to failure cannot be observed. In Example 2.9, no information is given about the distribution of failures times after the first 30 minutes. In Example 2.10, the average run time was only 38 minutes, and most AFW missions lasted for less than one hour. In such cases the exponential distribution, restricted to the observed time period, is a simple, reasonable approximation of the observable portion of the distribution.

Two assumptions are made concerning the physical process:

1. Assuming that no failure has occurred by time t, the probability that a failure will occur in a short time period t to $t + \Delta t$ is approximately proportional to the length of the exposure period, Δt. The probability does not depend on the starting time of the period, t, or on anything else.

2. Failures of distinct systems, or of one system during distinct missions, are independent of each other.

The kind of observable data is as follows:

- For each observed mission, the run time is observable. Also, it is known whether the run terminated in failure or in successful completion of the mission. As a minimum, the total run time and the number of failures to run are observed.

Assumption 1 implies that the time to failure is exponentially distributed with parameter θ. The interpretation of θ is that if the system is running, the probability of failure in the next short interval of length Δt is approximately $\theta \Delta t$. That is

$$\theta \Delta t \approx \Pr(t < T \le t + \Delta t \mid T > t),$$

where T is the random time until failure. When defined this way, θ is sometimes called the **failure rate**, or **rate of failure to run**. Many authors use the term **hazard rate**, denoted by h, and discussed in Appendix A.4.4. Note, the definition of θ is different for repairable systems (Section 2.2) and nonrepairable systems (the present section), even though it is represented by the same Greek letter and is called "failure rate" in both cases. See Thompson (1981) for a reasonably clear discussion of the subtle differences, and the glossary of this handbook for a summary of the definitions. The topic is discussed further in Appendix A.4.4.

It is instructive to compare the models for failure to run and standby failure. The physical process is essentially identical, but the observable data differs in the two models. That is, Assumptions 1 and 2 in the two sections agree except for small differences of wording. However, the time of failure to run is observable, whereas the time of transition to a standby failure is never known.

It may also be somewhat instructive to compare the Assumptions 1 and 2 here with the Assumptions 1-3 of the Poisson process in Section 2.2.2. For the standby-

failure model and the failure-to-run model, Assumptions 1 and 2 do not explicitly include an assumption ruling out simultaneous failures. The reason is that simultaneous failures are ruled out by the other two assumptions: it is not meaningful for a system to fail twice simultaneously; and distinct systems are assumed to fail independently of each other, and therefore not exactly simultaneously.

2.4.3 Data Needed to Validate the Model and Estimate θ

Suppose that the time to failure has an exponential distribution. Then, any reasonable estimator of θ needs only two pieces of information: the total running time, t, in the data period, and the number of failures to run, x, that occurred then.

However, more information is needed to investigate whether the exponential distribution is valid. Assumption 1 says that θ is constant during the mission. To investigate this, the analyst should know the failure times, that is, how long the failed pumps ran before failing. The analyst should also know the mission times, that is, how long the system ran when it did not fail; often, however, this information is not recorded and can only be estimated or approximated.

Implicit in Assumption 1 is that θ is the same over all the years of data, at all the plants where the data were collected. To investigate this, the data should be divided into subsets, corresponding to the different plants and years. Then the failure count and running time, x_i and t_i, should be given for each subset. This is the exact analogue of what was said in Section 2.2.3 for initiating events.

2.4.4 Case Studies: Validity of Model Assumptions in Examples

Consider now whether the assumptions of the model are plausible for the two examples.

Example 2.9 EDG Failures to Run

Assumption 1 says that a running EDG is as likely to fail in one short time interval as in any other time interval of the same length. That is, the EDG does not experience burn-in or wear-out failures. The reference report (Grant et al. 1999b) says that this is not true over a 24-hr mission. Indeed, that report divides the EDG mission into three time periods (first half hour, from one-half hour to 14 hours, and from 14 to 24 hours) to account for different failure rates during different time periods. Within the first half hour, however, the data do not give reason for believing that any short time interval is more likely to have a failure than any other time interval. Therefore, Assumption 1 can be accepted.

Assumption 2 is violated by common-cause failures. It is also violated if a failure's root cause is incorrectly diagnosed, and persists on the next demand. If these two conditions are rare the assumption may be an adequate approximation. More subtle dependencies are difficult to detect from data.

Example 2.10 AFW Turbine Train Failures to Run

Assumption 1 says that a running turbine train is as likely to fail in one short time interval as in any other time interval of the same length. The data are too sparse – only 2 observed failures – to confirm or refute this assumption. The data are also too sparse to confirm or refute Assumption 2, although failures in separate plants are virtually certain to be independent. In such a situation, it is common to accept the simple model as adequate. A more complicated model is justified only when a larger data set is available.

2.4.5 Discussion

The exponential time to failure can also be derived as the time to *first* failure in a Poisson process of Section 2.2. This is possible because the time to first failure and the times between subsequent failures are all exponentially distributed when the failures follow a Poisson process. The present context is simpler, however, because the process ends after the first event, failure to run. The Poisson-process assumptions about hypothetical additional failures are irrelevant.

2.5 Recovery Times and Other Random Duration Times

This section is about modeling of time data. Often, a measurement of interest is a random duration time, such as the time required to return a failed system to service or the lifetime of a piece of hardware. The distinction between random duration times here and events in time in Sections 2.2 and 2.4 is that here the individual times are measured on a continuous scale with units such as minutes or hours, while the earlier data sets involve discrete counts of the number of events occurring in a total length of time.

2.5.1 Examples

Here are some examples involving random duration times. They are only summarized here. Actual examples, with lists of durations times, will be analyzed in Chapter 6.

Example 2.11 Recovery times from loss of offsite power

A plant occasionally loses offsite power. When this happens, the plant reports the time until power is restored. Atwood et al. (1998) present such durations for LOSP events in 1980-1996.

Example 2.12 Repair times for turbine-driven pumps

A turbine-driven pump must occasionally be taken out of service for unplanned maintenance. The duration of time out of service for maintenance may be extractable from maintenance records.

Example 2.13 Time to failure of a component

A typical power plant will have many individual components such as compressors. When a component is put into service, it operates intermittently until it fails to perform its required function for some reason. Høyland and Rausand (1994) give an example of such data.

Example 2.14 Times to suppress fires

When a fire occurs in a nuclear power plant, the time until the fire is suppressed is of interest. Nowlen et al. (2002) report on analysis of such suppression times. One difficulty is that the time of fire onset often is not exactly known.

Example 2.15 Gradual degradation until failure

Examples 2.7 (steam binding) and 2.8 (failure of isolation valves) involve gradual degradation, which builds up until the system is inoperable. The time until the system is inoperable can be modeled as a duration time.

The common element in these examples is a duration time that varies in an unpredictable way. In Examples 2.11 and 2.12, the recovery time is composed of several factors such as the time to diagnose, perform and test repairs, and the time to complete documentation required before returning the plant to normal operating conditions. Example 2.13 is a failure-to-run example, similar to those of Section 2.4. This example differs from that of Section 2.4, however, because here it is assumed that virtually all of the times to failure are recorded. In Section 2.4, on the other hand, most of the systems did not fail during the test period or operational mission. The severe truncation of the data in Section 2.4 meant that only a simple model could be considered. The more complete data here allows analysis of a more complex model. Example 2.14 is complicated by the lack of exact knowledge of the duration time. Finally, Example 2.15 gives a realistic conceptual way to model the gradual degradations encountered in Section 2.3.1, although good data are unobtainable.

All five examples involve a duration time that is uncertain due to random factors. Consequently, the duration times are modeled as continuous random variables.

2.5.2 Duration-Time Models

The duration, T, is random, following some probability distribution. Two assumptions are made about the process:

1. Each duration is statistically independent of the others, and

2. All the random durations come from the same probability distribution.

The data description is simple:

- The individual durations are observable. As a bare minimum, the number of durations and the total duration time are observed.

Assumptions 1 and 2 can be summarized by saying that the durations are **independent** and **identically distributed**. Independence means that one duration does not influence the probability of any other duration. The assumption of identical distributions means that each random duration is as likely as any other to be long or short. If the durations are from distinct systems, the systems are assumed to be identical and to act independently. If the durations are in sequence, as for a system that alternates being up and down, the assumption implies that no learning or long-term aging takes place, and that each repair restores the system to a condition as good as new. Such a process is called a **renewal process**.

The assumptions do not require a particular distribution for the time between events. The most important such distributions in PRA applications are:

- lognormal,
- exponential,
- Weibull, and
- gamma.

These distributions are summarized in Appendix A.7. An important part of the data analysis consists of deciding on the form (or several plausible forms) of the distribution. This will be discussed in Chapter 6. For now, we simply note that these and other distributions are possible.

There are different ways to specify a probability distribution, and the next material summarizes some of the concepts: their definitions, how to interpret them, and how they are related to each other. The data-analysis techniques of Chapter 6 will use these ways of characterizing distributions. The usual convention is to denote the random variables using capital letters, *T*, and observed times as lower case, *t*. The letter *T* is used, rather than some other letter such as *X*, because the random quantities are times. As seen from the examples, the durations may be times to repair, times to failure, or other times. However, the concepts and formulas are valid for any application.

The **cumulative distribution function** (c.d.f.) of a real-valued random variable *T* is defined as

$F(t) = \Pr(T \# t)$

for all real numbers *t*. The name is sometimes abbreviated to **distribution function**. The c.d.f. is the probability that the random variable *T* will assume a value that is less than or equal to *t*. The c.d.f. is a monotonically increasing function of *t*, with the limiting properties $F(0) = 0$ and $F(+4) = 1$. [For random variables that, unlike durations, can take negative values, the limiting properties are $F(!4) = 0$ and $F(+4) = 1$. That general case has few applications in this handbook.]

The distribution is commonly used to characterize the lifetimes, or recovery times, or some other kind of durations, of a whole population of systems. The population might be a large set of identical systems that are operating in similar applications and with durations that vary due to random influences. $F(t)$ is the fraction of items that have durations *t* or less, in a hypothetical infinite population.

A related function, denoted by $f(t)$, is called a **probability density function** (p.d.f.) for a continuously distributed positive-valued random variable *T*. It is related to the c.d.f. by

$$f(t) = \frac{d}{dt} F(t) \quad \text{and}$$

$$F(t) = \int_0^t f(u)du \quad .$$

The variable *u* is a dummy variable of integration, and *t* is the upper limit of the integral. An example of a p.d.f. and the associated c.d.f. are shown in Figure 2.1.

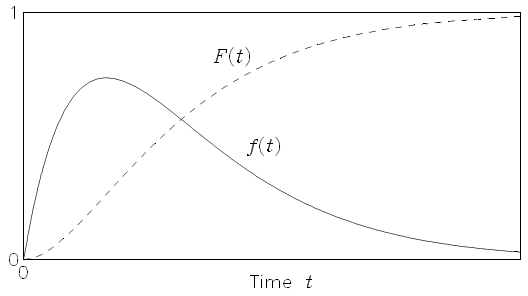


Figure 2.1 Probability density function (p.d.f.) and cumulative distribution function (c.d.f.).

It follows that probabilities corresponding to occurrences in a small interval of time are approximately proportional to the p.d.f.,

$\Pr(t < T \# t +)t) \ . \ f(t)) t.$

Therefore, the ordinate of a p.d.f. has units of "probability density" and not probability (as for a c.d.f.). Thus, a p.d.f. determines how to assign probability over small intervals of time. Now consider an arbitrary interval from *a* to *b*. In this case we have

$$\Pr(a < T \leq b) = \int_a^b f(t)dt \quad .$$

The simplest distribution is the **exponential** distribution. It arises when Assumption 1 of Section 2.4.2 is satisfied. (That assumption is phrased as if *T* is a time until failure.) In that case, the probability distribution is exponential, and determined by a single parameter, θ. The p.d.f. and c.d.f. are given by

$f(t) = \theta e^{!\theta t}$

$F(t) = 1 ! e^{!\theta t}$. (2.6)

When deriving the distribution mathematically from Assumption 1, it is necessary to assume that $F(0) = 0$, that is, failures at time 0 have zero probability. Although not stated explicitly, this assumption is implicit in the context of failure to run, because any failures at time 0 would be counted as failures to start, not failures to run.

2.5.3 Data Needed to Estimate Distribution of Durations and Validate Model

In general, a sample of observed durations is needed to estimate the distribution of duration times. These durations must independent and identically distributed, that is, they must be generated by a process satisfying the two assumptions given at the beginning of Section 2.5.2.

The special case when the times are assumed to have an exponential (δ) distribution is simpler. Only the number of durations and the total duration time are needed to estimate δ,. However, the individual durations are still needed to investigate whether the distribution is exponential or of some other form. Incidentally, when the distribution is assumed to be exponential, the model given here differs from the standby-failure model (Section 2.3.3.1.1) and from the failure-to-run model (Section 2.4.2) *only* by the kind of data that can be observed.

To validate whether the distribution is the same for all the data, extra information should be recorded for each duration, the relevant circumstances of each duration. The circumstances of interest are those that might affect the durations, such as time of the event, system location, and system condition just before the event.

2.5.4 Case Studies: Validity of Model Assumptions in the Examples

Examples 2.11 through 2.13 all appear to satisfy the assumptions of Section 2.5.2. Example 2.14 also does, except that the durations are not observed exactly.

In each case, all the distributions come from some distribution. Discovering the form of that distribution is a task for the data analyst.

One might ask whether the durations are statistically independent. For example, does a long repair time for a turbine-driven pump add an extra benefit to the pump, so that the next few repair times will be short?

One might also ask, for each example, whether the durations all come from the same probability distribution. For example, if the data cover a period of years, has there been any long-term learning, so that recovery times or repair times tend to be shorter than at the start of the data period? Are different durations associated with different systems for the turbine-driven pumps, with different causes of loss of offsite power, or with different kinds of fires?

The above are questions that could be investigated during the data analysis, if enough durations have been observed.

Example 2.14 is complicated by lack of exact measurements of the durations. Bounds can be given, and the analysis must be based on these upper and lower bounds rather than on exact times.

Example 2.15 is different because the durations are not observable at all. It might be theoretically interesting to model the time until the system is in a failed condition as a duration, but there is no monitor on the pump or valve that says, “At this time the system just became inoperable.” Therefore, the durations are not directly observable, not even in principle. Therefore, the methods of this handbook are not applicable to this example.

Fortunately, degradation mechanisms have become minor contributors to risk. When a degradation mechanism is recognized as important, the natural response is not to collect data to better estimate the rate of degradation. Instead, the natural response is (a) to shorten the interval between preventive maintenance activities, and so to identify and correct incipient degradation, or (b) to modify the plant to mitigate or eliminate the problem. Examples are the apparent elimination of steam-binding in AFW pumps, mentioned above, and of intergranular stress corrosion cracking (IGSCC) in BWR piping (Poloski et al. 1999a, Appendix J).

2.6 Unavailability

This section considers **test-and-maintenance unavailability**, corresponding to intentional removal of the equipment from service for testing and/or maintenance. This section does not consider unavailability resulting from the hardware being in an unrecognized failed condition; that topic was treated in Section 2.3.3.

The discussion here is presented in terms of trains, although other hardware configurations, such as individual components, could be considered equally well. A standby train, such as the single train of the HPCI system or a motor-driven train of the AFW system, is normally available if it should be demanded,

but sometimes it is out of service for planned or unplanned maintenance. The event of a train being unavailable is called an **outage**, and the length of time when it is unavailable is called an **outage time** or **out-of-service time**. In a data set, the **exposure time** is the time (e.g. number of hours) when the train should have been available. The **unavailability** is the long-term ratio of outage time to exposure time – the fraction of time that the system is out of service when it should be available. More precisely, the **planned-maintenance unavailability** is the fraction of time that the system is out of service for planned testing and maintenance, and the **unplanned-maintenance unavailability** is defined similarly. In summary, outage times are random but the unavailability is a parameter, an unknown constant, denoted here by q. Subscripts such as "planned" and "unplanned" can be attached to q for clarity if needed.

2.6.1 Example

Example 2.16 CVC unavailability for test and maintenance

Train outages of various durations occurred during 15 calendar months at a plant with two trains in the chemical and volume control (CVC) system. For each month, the outage durations are given by Atwood and Engelhardt (2003).

A way to picture the status of a standby train or other repairable system uses a **state variable**, defined as $S(t) = 1$ if the system is up at time t, and $S(t) = 0$ if it is down at time t. A particular system history is illustrated in Figure 2.2, from Engelhardt (1996). This figure shows when a particular system was operating ($S = 1$) or shut down ($S = 0$). A nominally identical system would have a somewhat different history for the same period, or the same system would have a different history over a different time period of the same length.

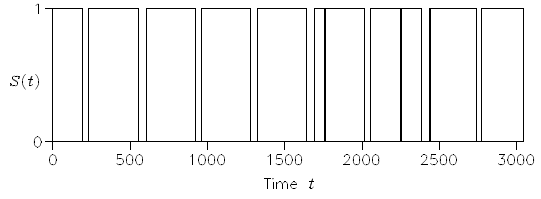


Figure 2.2 Uptime and downtime status for one system.

As stated above, the long-term fraction of time when the system is down is called the system **unavailability**.

2.6.2 Probability Model

The assumed underlying model is an **alternating renewal process**. At any point in time a system is in one of two states: "up" or "down," corresponding in our application to being available or out of service. Initially, the system is up, and it remains up for a random time Y_1; it then goes down, and stays down for a random time Z_1. Then it then goes up for a time Y_2, and then down for a time Z_2, and so forth. The assumptions needed for the data analysis methods in Chapter 6 are the following:

1. The random variables Y_i have one distribution that is continuous with a finite mean, and so do the random variables Z_i.
2. All the random variables are independent of each other

The sum of the down times, $\sum Z_i$, is the total outage time in the data. The sum of all the times, $\sum Y_i + \sum Z_i$, is the exposure time – the time when the system should be available. Time when the system is not required to be available is not counted in either the up time or the down time.

Two kinds of data can be considered:

- **Detailed data**: the onset time and duration of each individual outage are recorded, as well as the total time when the train should have been available; and

- **Summary data.** Data totals are given for "reporting periods," such as calendar months. For each reporting period, the total outage time and exposure time are recorded.

Section 6.7 describes how to analyze both types of data.

2.6.3 Data Needed to Validate the Model and Estimate q

The unavailability, q, can be estimated from either kind of data. Enough data should be collected so that any periodic, lengthy, planned outages are appropriately represented – neither over-represented nor under-represented.

In addition, if summary data are used, the methods given in Chapter 6 combine reporting periods into larger subsets of the data, at the very least so that the aggregated subsets do not contain outage times of zero. Therefore, a large enough set of summary data is

needed so that it consists of at least two (as a bare minimum) subsets of approximately equal exposure time, with each subset containing nonzero outage time.

To validate the model, any information that might be related to unavailability should be recorded. For example, if a motor-driven pump has most of its scheduled maintenance during the plant's refueling outages, and the pump's availability during shutdown is of interest, then the data should indicate which outages and exposure times correspond to reactor shutdown. Separate analyses will probably need to be performed for the time when the reactor is up and when the reactor is down, to keep Assumption 1 from being violated.

2.6.4 Case Study: Validity of Model Assumptions in Example

The ideas here are applicable to virtually any system, with Example 2.16 being just one example.

The trains may undergo periodic, infrequent, lengthy testing and maintenance, and less lengthy testing and maintenance at more frequent intervals. This periodicity of planned maintenance means that Assumption 2 cannot be exactly true. The lengthiest outages tend to be evenly spaced, not random as assumed. However, more realistic assumptions would be very difficult to work with.

It seems plausible that this deterministic periodicity should lead to conservative estimates. That is, analysis methods that assume pure randomness will tend to overestimate the variance, so that the resulting uncertainty in q is overestimated. However, this conjecture has not been carefully investigated, and the 15 months of data in Example 2.16, analyzed in Section 6.7, do not support the conjecture.

Assumption 1, on the other hand, is surely correct. The distributions are continuous, and it is inconceivable that the durations for an operating power plant would have infinite means.

3. COMPONENT FAILURE AND BOUNDARY DEFINITIONS

3.1 Failure Definitions

While the terms "faults" and "failures" are casually used interchangeably, in the context of fault tree analysis these terms have more distinct meanings. Thus, for data analysis, it is necessary for one to understand the distinctions. Generally speaking, all failures are faults, but not all faults are failures. To put it another way, failures comprise a subset of the larger set of faults. For probabilistic risk assessment (PRA) purposes, **failures** are regarded as basic (and undesired) events which:

- render a component, subsystem, or system incapable of performing its intended function,
- represents a basic fault tree input that is not analyzed further, and
- require numerical estimates if quantification is to be performed.

Faults, on the other hand, are higher order events (representing the occurrence or existence of an undesired state of a component or set of components) which are analyzed further, and ultimately resolved into their constituent failures (Breeding, Leahy, and Young 1985; ANS and IEEE 1983; and Vesely et al. 1981).

The failures modeled in PRA can have many causes or mechanisms. For example, failure of a motor-operated valve (MOV) to open on demand can occur due to physical problems with the valve (stem failure, disc separation, etc.), problems with the motor operator (motor failure, control circuit failure, breaker failure, etc.), or due to loss of motive or control power. In addition, the MOV may be unavailable due to test or maintenance on its constituent parts. As such, each failure (i.e., basic event) is the sum of the contributions from each piece-part included in the component boundary. Thus, it is critical to define what the component boundary is in order to get the right data.

3.2 Component Boundary Definitions

In order to collect failure data for components, it is necessary to define **component boundaries** by specifying the scope of each item to be considered as a single entity. The PRA model and the data collection should be coordinated so that the boundaries of the components are defined identically. For example, all pieces of an MOV are typically considered to be part of a single "component" when collecting reliability data even though the valve consists of various piece parts (e.g., electric motor, gearbox, limit switches, torque switches, reversing contacts and coils, stem, disc, valve body, etc.) that may be separately identified in the plant maintenance records. PRAs typically do not model failures of every switch, relay, or contact in a control circuit of a pump because that type of detail is difficult to obtain from the plant data. Instead, failures of these components are typically included with actual failures of the pump to establish a pump failure rate.

If generic data sources are used, it becomes the responsibility of the analyst to ensure that the component boundary definitions used in the generic data source are compatible with the boundary definitions used by the PRA being performed.

Some typical examples of component boundaries are shown in Table 3.l. The boundaries of a component should include all components specific to the component. However, the component boundary should not include piece-parts that are shared with other components modeled in the PRA. For example, the component boundary for emergency-actuated valves commonly includes the valve control circuit. However, the components needed to generate an actuation signal that initiates multiple components modeled in the PRA should not be included as part of that specific valve boundary. Similarly, a diesel generator boundary will typically include the fuel day tank but the fuel oil transfer pumps are not included since they are required for operation of all the plant's diesel generators.

3.3 Failure Severity

The raw data for a specific component will contain some events not relevant to the component failure modes being analyzed. These events can be screened from further analysis. Some of the events will be component failures that should be included in the data assessment. The type of component failures will determine how they are classified and subsequently used to generate the required component failure data.

Component malfunction events are commonly classified into one of the following three **failure severity** categories:

Table 3.1 Examples of component boundaries.

Component	Component Boundary
Diesel Generators	The diesel generator boundary includes the generator body, generator actuator, lubrication system (local), fuel system (local), cooling components (local), startup air system, exhaust and combustion air system, individual diesel generator control system, circuit breaker for supply to safeguard buses and their associated local control circuit (coil, auxiliary contacts, wiring, and control circuit contacts) with the exception of all the contacts and relays which interact with other electrical or control systems.
Motor Pumps	The pump boundary includes the pump body, motor/actuator, lubrication system cooling components of the pump seals, the voltage supply breaker, and its associated local control circuit (coil, auxiliary contacts, wiring, and control circuit contacts).
Turbine-Driven Pumps	The turbine-driven pump boundary includes the pump body, turbine/actuator, lubrication system (including pump), extractions, turbopump seal, cooling components, and local turbine control system (speed).
Motor-Operated Valves	The valve boundary inc1udes the valve body, motor/actuator, the voltage supply breaker and its associated local open/close circuit (open/close switches, auxiliary and switch contacts, and wiring and switch energization contacts).
Air-Operated Valves	The valve boundary includes the valve body, the air operator, associated solenoid-operated valve, the power supply breaker or fuse for the solenoid valve, and its associated control circuit (open/close switches and local auxiliary and switch contacts).
Fans	The fan boundary includes the fan, the voltage supply breaker, and its associated control circuit (open/close switches and local auxiliary and switch contacts).
Batteries	The battery component boundary typically includes just the battery. Battery chargers are modeled as separate components.
Bus Circuit Breakers	A bus circuit breaker boundary includes the breaker and its associated control circuit (open/close switches and local auxiliary and switch contacts).

- catastrophic failures,
- degraded failures, and
- incipient failures.

A **catastrophic (complete) failure** is one that prevents the component from performing its mission as defined in the PRA (Whitehead 1993). Catastrophic failures require some kind of repair or replacement action on the component in order to restore the component to operability. For example, a valve that fails to open due to a valve operator mechanical failure is a catastrophic failure.

A **degraded failure** is such that a component can perform its mission, but at less than the optimum performance level (Whitehead et al. 1993). An **incipient failure** is such that there is no significant degradation in performance but there are indications of a developing fault (Whitehead et al. 1993). The difference between the two is generally a matter of severity. For example, an event involving pump shaft vibration indicates possible damage to the pump bearings. Severe vibration may be considered as degraded failure if the pump produces less than maximum flow. Shaft seizure or other failures could occur within a few hours if the pump remains running and thus would likely be removed from operation for corrective maintenance. In contrast, minor vibration may not result in degraded flow. This would thus be an incipient failure. The significance of this event is that it also could result in removal of the pump from operation for inspection, lubrication, or some other corrective action. Information about the types of repairs made, the parts replaced, and the urgency of the repairs often provides important insight about the severity of these two types of component failures.

Although both degraded and incipient failures will typically lead to a corrective action, the corrective action may or may not make the component unavailable to perform its function. For example, maintenance on the operator of a valve that is normally open will not lead to the unavailability of the valve if it is required to be open for system operation. This illustrates the importance of ascertaining from event records the modes of a component operation that a corrective action would prevent.

Sometimes the event information is so unclear and incomplete that a definite classification of the severity of a component malfunction event is not possible. For example, Mosleh and Apostolakis (1985) cites one maintenance work request issued at a nuclear power plant that described the problem as follows: "Check valve RHR-V-1A is leaking badly." The maintenance foreman's description of the corrective action read: "Fixed it, not leaking anymore!" No further information was available. From the description given, one cannot say for sure whether the leak was internal or external, or whether it was large enough to result in functional failure of the check valve.

Unfortunately, the above example is not uncommon. Descriptions of the malfunctions and repairs are often very brief. The data analyst, then, is faced with the difficult task of deciding whether to call a malfunction a failure or not. The inability to distinguish between severity levels of failures is particularly important as the difference between the frequencies of catastrophic and degraded modes of failures can be significant. Therefore, in the absence of sufficient information, the conservative assumption could be made that all such events be recorded as catastrophic failures. Unfortunately, conservative categorization of uncertain events can lead to significantly higher failure rates.

Ultimately, the definition of failure from the system analysis decides the classification of the data. Thus, the failure of a component must match the definition of the failure as described in the PRA model. A component must fail to perform its function as defined in the model. For example, a relief valve that opens at 1,115 psig instead of the required 1,110 psig is not failed, although it may be described as failed by the governing technical specifications, and a pump that delivers 645 gpm instead of the required 700 gpm is not failed if 645 gpm is sufficient for the function that it is required to perform.

4. DATA SOURCES

Two types of data sources can be utilized to produce the various parameter estimates that are needed in a probabilistic risk assessment (PRA). This chapter identifies and discusses these two data sources. Section 4.1 identifies and discusses plant-specific data sources. Section 4.2 does the same for generic data sources.

4.1 Plant-Specific Data Sources

Use of plant-specific data in a PRA produces risk estimates that reflect the actual plant experience.

The scope of a plant-specific data analysis is determined by the events that are included in the PRA models. In general, plant-specific data are generally reviewed for the following types of events:

1. The accident initiating events analyzed in the PRA.

2. The components included in system models (generally fault trees). For components the definition includes the component boundary and failure mode. For unavailabilities due to maintenance or testing it is necessary to know whether the unavailabilities are to be specified at the component, segment, train, or system level.

3. Some recovery events included in the PRA models. Although most recovery events are analyzed using human reliability analysis, the probabilities of some events can be based upon a review of operating experience.

Once the data needs are identified, the sources of raw data at the plant are identified. In most cases, the information needed may have to come from multiple sources. For example, identification of maintenance events and their duration may come from a control room log, but other sources such as maintenance work requests may be required to determine other information such as whether a component had experienced a catastrophic or degraded failure.

There are many sources of raw data at a nuclear power plant. Different plants have different means of recording information on initiating events and component failure and maintenance events. Since no one source exists at a nuclear power plant that contains all the necessary data, different sources must be reviewed. The ease in which the plant-specific data can be interpreted and the subsequent quality of the resulting parameter estimates are a function of how well the plant personnel recorded the necessary information.

Basic requirements associated with raw data sources and some typical sources of raw data available at nuclear power plants are identified in the following sections.

4.1.1 Requirements on Data Sources

There are a variety of data sources that exist at a plant and can be used in a data analysis. However, there are some basic requirements that these raw data sources should meet in order to be useful. Some typical requirements, some of which were suggested in EPRI TR-100381 (EPRI 1992), are delineated below.

4.1.1.1 Initiating Events

For reports on initiating events it is essential to include the status of those systems that would be impacted as a result of the event. This is typically not a problem since the Licensee Event Report (LER) that is required to be filed with the Nuclear Regulatory Commission (NRC) following a plant trip usually contains this type of information. It is also common for utilities to generate additional detailed trip reports that delineate the cause and effects of the event. Such reports need to specify critical information needed for data analysis such as the power level at the time of the plant trip and the sequence of events, including the timing of individual events.

4.1.1.2 Component Failures

For each event at a plant resulting in the unavailability of a component, it is necessary that the raw data sources identify the particular component or set of components associated with the event. In order to determine if a specific event contributes to a particular component failure mode or to an unavailability due to the component being in maintenance (either preventive or corrective), it is necessary to be able to distinguish between different degrees of degradation or failure. The event reports should therefore specify whether maintenance was required and if the maintenance was corrective or preventive. If the component maintenance is preventive there is generally no failure that initiates the maintenance.

If an event involves corrective maintenance, information is required to allow determination of the severity of the failure (see Section 3.3 for definitions of event severity). The ability to distinguish between severity levels of failures is particularly important since the difference between the frequencies of catastrophic and degraded modes of failures can be significant. In addition, information is required to determine the component in which the failure actually occurred and the mode of failure. Finally, it should be possible to determine the time the component is unavailable during each maintenance event.

The data analysis may use plant data on component unavailability that is being collected for other than PRA purposes. The requirements for recording the data for these other purposes may use definitions of severity and failure modes that are different from the PRA definitions. The definitions used for the data collection programs should be determined and an appropriate translation to the PRA basic events made.

4.1.1.3 Recovery Events

The information needed to estimate the probabilities associated with recovering specific components or systems from a failed state is similar to that needed for component failures. Specific information pertaining to the type of failure experienced by the component or system (e.g., fail to operate, fail to start, fail to run), the number of repair occurrences, and the time required to perform the repair is needed to produce component repair probabilities.

4.1.2 Data Sources

Data sources that can provide information for determining the number of initiating events include:

- internal plant failure records (e.g., scram reports or unusual event reports),
- operator logs,
- LERs, and
- monthly operating reports/Gray Book.

Some data sources that typically provide information on the occurrence of component failures include:

- LERs,
- internal plant failure records (e.g., failure reports, trouble reports, or unusual event reports),
- maintenance records (e.g., maintenance work orders, work request records),
- plant logs (e.g., control room log, component history logs), and
- data bases (e.g., Equipment Performance and Information Exchange System/Nuclear Plant Reliability Data System).

The evaluation of component failure rates also requires the number of demands and operating time for the components. Sources of data for these parameters include:

- monthly operating reports/Gray Book,
- component history logs,
- plant population lists,
- test procedures,
- plant operating procedures, and
- component demand or operating time counters

Repair information can be obtained from sources such as:

- plant logs and
- maintenance work orders.

The type of information available in these sources and their limitations are discussed in the following sections.

4.1.2.1 Regulatory Reports

All plants are required to submit LERs to the NRC for all events meeting the 10 CFR 50.73 reporting criteria presented in NUREG-1022 (NRC 2000a). LERs deal with significant events related to the plant, including plant shutdowns required by the technical specifications, multiple train failures, engineered safety feature actuations, and conditions outside the design basis or not covered by plant procedures. An LER includes an abstract that describes the major occurrences during the event; the components, systems, or human failures that contributed to the event; the failure mode, mechanism, and effect of each failed component; and an estimate of the elapsed time from the discovery of the failure until the safety system train was returned to service. A computerized search of LER information is possible using the Sequence Coding and Search System (SCSS) (Gallaher et al. 1984).

LERs generally provide a good description of the causes of a reactor trip and subsequent events. However, their value for obtaining component failure data is very limited. The reporting criteria are limited to safety-related trains or system failures, and therefore LERs are not generally submitted for all failures. Furthermore, LERs may not be submitted for every

safety-related component failure since individual component failures do not have to be reported if redundant equipment in the same system was operable and available to perform the safety function. The reporting criteria for LERs are also subject to interpretation by the persons generating the reports and thus can lead to inconsistencies in the LER data base. Furthermore, there are other perceived deficiencies in the LERs (Whitehead et al. 1993) that limit the usefulness of the LER system for use in obtaining estimates of component failure rates. The NRC staff prepared NUREG-1022, Revision 1 (NRC 1998), to address general issues in reporting that have not been consistently applied. It covers some of the issues identified above.

The LER rule published in 1983 has recently been amended and the reporting guidance in NUREG-1022, Revision 2 (NRC 2000a) has been revised to eliminate the burden of reporting events of little or no safety significance, to better align the rules with the NRC's current needs and to clarify the reporting guidance where needed. However, the rule still only requires the reporting of failures leading to the unavailability of safety-related system trains. Thus, LERs will not provide failure data for all risk significant components.

In summary, LERs are a good source for identifying and grouping initiating events. However, they have very limited value for obtaining component failure data.

A plant's Technical Specifications requires that a monthly operating report be provided by the plant licensee to the NRC. The scope of the information requested of the licensees was originally identified in Draft Regulatory Guide 1.16 (NRC 1975a) and includes operating statistics and shutdown experience information. The information requested to be included in the monthly operating report contents was revised by Generic Letter 97-02 (NRC 1997) and eliminated some reporting requirements. Information that still must be reported includes identification of all plant shutdowns, whether they were forced or scheduled shutdowns, their duration, the reason for the shutdown, the method of shutting down the reactor, and corrective actions that were taken. In addition, the monthly operating reports include the number of hours the reactor was critical, the number of hours the generator was on line, and the net electrical output of the plant.

The NRC initially compiled the information from the monthly operating reports on a monthly basis and published it in a hard copy form as NUREG-0020, "Licensed Operating Reactors - Status Summary Report" (NRC 1995b). This document is referred to as the "Gray Book." NUREG-0020 was discontinued after the December 1995 report. However, the data requested in Generic Letter 97-02 is being collected and computerized as part of the NRC Performance Indicator Project.

In summary, the monthly operating reports provide information on the number of scrams, the time spent at full power, and the time spent in shutdown. This information can be used in identifying and grouping initiating events and in calculating the exposure time in which they occurred. It is important to note that this same information is generally available from the control room logs and other sources. Thus, in general, the monthly operating reports can be used to supplement or verify other data sources.

4.1.2.2 Internal Plant Failure Reports

Different plants have different means of recording initiating events and component failures. For each automatic and manual scram, most plants generate an internal scram report. Scram reports generally cover the same information provided in LERs and monthly operating reports. Thus, they can be used as the primary or supplementary source for evaluating plant scrams.

Most plants have a means of recording component failures, records that are for the licensee's own use rather than for a regulatory use. Reports are generally created when significant component failures or degraded states occur during plant operation or are identified during plant surveillance tests. These reports may be called Unusual Occurrence Reports, Action Reports, Failure Reports, Discrepancy Reports, or Trouble Reports. Some of the events documented in these reports may lead to an LER. However, these reports may not identify all component failures and generally are not exhaustive. Thus, these reports are useful for supplemental information but are not a good source of component reliability data.

4.1.2.3 Maintenance Records

At all plants, some form of written authorization form is required to initiate corrective or preventative maintenance work, or design changes. These authorization forms are known under different names at various plants including work request/completion records, maintenance work orders, clearance requests, work requests, or tag-out orders. Maintenance records are a primary source of component failure data since

they usually identify the component being maintained, whether the component has failed or is degraded, the corrective action taken, and the duration of the maintenance action. The time of the failure is also available but maintenance records generally contain limited information on the impact, cause, and method of discovery of the component failure.

4.1.2.4 Plant Logs

At each plant, a control room log is typically completed for each shift and contains a record of all important events at a plant. Control room logs identify power level and plant mode changes, essential equipment status changes, major system and equipment tests, and entry and exit of Technical Specification Limiting Conditions of Operation (LCOs). When properly maintained, a control room log is a good source of information on major equipment and unit availability. However, the amount of information entered can vary from shift to shift. Furthermore, the entries tend to be brief.

The control room logs are difficult to use as a source of maintenance data since the tag-out and tag-in for a maintenance event may span days or even months and may not be dutifully recorded. The control room logs are also limited in value as a source of component failure data since not all failures may be recorded by the operators. Component maintenance and failure information is generally found more easily in maintenance work orders. All plant trips are likely to be recorded on control room logs, but likely will not include a description of the cause of the trip or the subsequent transient behavior. LERs or plant scram reports must be reviewed to obtain this additional information.

In summary, control room logs are good supplementary sources of information but there are usually more convenient and complete sources of information available such as maintenance records. However, the control room logs are probably the best source of data for indicating when redundant system trains are switched from operating to standby status.

There may be other logs at a plant that contain essential data. One example is a component history log. These logs typically contain data on every failure and maintenance and test action for a given component. As such, component history logs are good sources for identifying not only the number of component failures, but also the number of demands a component experiences.

4.1.2.5 Component Exposure Data Sources

Calculation of plant-specific failure rates requires determination of the number of failures and the corresponding number of demands or operating time. As indicated in the previous subsections, some of the data sources used to establish the number of failures also contain information on the number of demands and operating time. However, these sources do not contain all component demands or the operating time for all components. Additional documents that must be reviewed for information about component demands and operating hours include test procedures.

In addition to demands presented by automatic initiations and maintenance activities (obtained from sources such as control room logs and maintenance records), periodic testing is an important source of demands especially for safety-related equipment. To establish the number of demands due to testing, testing procedures pertinent to a component must be reviewed. In addition to the actual test demands, additional test demands may be imposed by technical specifications following failure of a component. A typical example where this is imposed is when a diesel generator is unavailable for operation. Test logs or similar records can be examined to obtain an estimate of the number of tests carried out during the time period of interest.

It should also be noted that at some plants, some major components may be monitored to count the number of actuations experienced by the breakers (breaker cycle counters). In addition, the operating hours for large motor-driven components at some plants may be automatically registered on running time meters at the electrical switchgear. Such counters and logs can be used to supplement the demand and operating time information obtained from other sources.

4.1.3 Plant-Specific Data Bases

The Institute of Nuclear Power Operations (INPO) has maintained several databases of component failure data provided by each nuclear power plant since 1984. The first, Nuclear Plant Reliability Data System (NPRDS), was a proprietary computer-based collection of engineering, operational, and failure data on systems and components in U.S. nuclear power plants through 1996. The second, the Equipment Performance and Information Exchange (EPIX) System, replaced NPRDS and includes data reported since 1987. Both data bases are discussed in the following sections.

4.1.3.1 Nuclear Plant Reliability Data System (NPRDS)

In the early 1970s, industry committees of the American National Standards Institute (ANSI) and the Edison Electric Institute (EEI) recognized the need for failure data on nuclear plant components. As a result, a data collection system was developed whose objective was to make available reliability statistics (e.g., failure rates, mean-time-between-failures, mean-time-to-restore) for safety related systems and components.

This system, the Nuclear Plant Reliability Data System (Tashjian 1982), was developed by Southwest Research Institute (SwRI). Plants began reporting data on a voluntary basis in 1974, and continued reporting to SwRI until 1982. In January 1982, the INPO assumed management responsibility for the system until reporting was terminated at the end of 1996.

Originally the scope of the NPRDS covered the systems and components classified by ANSI standards as Safety Class 1, 2, or 1E, with a few exceptions such as reactor vessel internals and spent fuel storage. However, later the scope was expanded to cover any system important to safety and any system for which a loss of function can initiate significant plant transients (Simard 1983). By the end of 1984, 86 nuclear power plant units were supplying detailed design data and failure reports on some 4,000 to 5,000 plant components from 30 systems (Simard 1985).

Data reported to NPRDS consisted of two kinds: engineering reports and failure reports. The engineering reports provided detailed design and operating characteristics for each reportable component. The failure reports provided information on each reportable component whenever the component was unable to perform its intended function. The same operational data contained in NUREG-0200 was also included in the system. The NPRDS failure reports provided to INPO were generally generated by plant licensees utilizing maintenance records such as maintenance work orders. These reports utilized a standard set of component boundaries and failure mode definitions.

4.1.3.1.1 Limitations in the Data Available from the NPRDS

Several issues regarding the quality and utility of the NPRDS data have been observed, including:

1. Input to NPRDS was discontinued on December 31, 1996.
2. The number of component demands is provided by estimation.
3. The exposure time is estimated.
4. The amount of time needed to repair components out for corrective maintenance is not provided.
5. Maintenance rates are not provided.
6. The voluntary nature of the reporting system introduces uncertainty into measuring the frequency at which a particular type of problem occurs.
7. The final results of a problem investigation or the ultimate corrective action taken are not always included.
8. Report entries tend to be brief and often do not provide enough information to identify the exact failure mechanism.

4.1.3.2 Equipment Performance and Information Exchange (EPIX) System

The need for high-quality, plant-specific reliability and availability information to support risk-informed applications was one impetus for a proposed reliability data rule by the NRC to require utilities to provide such information. Instead of a regulatory rule, the nuclear industry committed to voluntarily report reliability information for risk-significant systems and equipment to the EPIX system. EPIX is a web-based database of component engineering and failure data developed by INPO to replace NPRDS. The utilities began reporting to EPIX on January 1, 1997.

EPIX enables sharing of engineering and failure information on selected components within the scope of the NRC's Maintenance Rule (10 CFR 50.65) and on equipment failures that cause power reductions. It also provides failure rate and reliability information for a limited number of risk-significant plant components. This includes components in the systems included in the scope of the Safety System Performance Indicator (SSPI) program. EPIX consists of:

- a site-specific database controlled by each INPO member site with web-based data entry and retrieval,
- an industry database on the INPO web site where selected parts of the site-specific database are shared among plants, and
- a retrieval tool that provides access to the vast historical equipment performance information available in the NPRDS.

Events reported to EPIX include both complete failures of components and degraded component operation. The number of demands and operating hours (i.e., reliability data) and the unavailability are required to be collected for key components in the SSPI safety systems for each plant. In addition, contributors to EPIX are also to include one-time estimates of the number of demands and run hours for other risk-significant components not included in SSPI systems.

4.1.3.3 Reliability and Availability Data System (RADS)

The NRC has developed the Reliability and Availability Data System (RADS) to provide the reliability and availability data needed by the NRC to perform generic and plant-specific assessments and to support PRA and risk-informed regulatory applications. The NRC is incorporating data from EPIX and INPO's SSPI system along with information from other data sources (e.g., LERs and monthly operating reports) into RADS. Data are available for the major components in the most risk-important systems in both boiling water reactors (BWRs) and pressurized water reactors (PWRs).

The reliability parameters that can be estimated using RADS are:

- probability of failure on demand,
- failure rate during operation (used to calculate probability of failure to continue operation),
- maintenance out-of-service unavailability (planned and unplanned), and
- time trends in reliability parameters.

The statistical methods available in RADS include classical statistical methods (maximum likelihood estimates and confidence intervals), Bayesian methods, tests for homogeneity of the data for deciding whether to pool the data or not, Empirical Bayes methods, and methods for trending the reliability parameters over time.

4.2 Generic Data Sources

Several generic data sources currently available and used throughout the nuclear power PRA industry are identified in this section. Several of these data bases are discussed with regard to their attributes, strengths, and weaknesses. Data bases for both initiating events and component failure rates are included. Some data sources represent compilations of raw data which have been collected directly from various facilities and processed and statistically analyzed. Other data sources utilize the results of the statistical analyzes of other data bases to derive estimates for component probabilities.

Section 4.2.1 contains discussions and summaries of generic data bases sponsored by the NRC for use in both government and industry PRAs. Section 4.2.2 contains discussions and summaries of generic data bases sponsored by the Department of Energy (DOE) for use in PRAs. Section 4.2.3 contains discussions and summaries of generic data bases developed by nuclear power industry related organizations. Section 4.2.4 contains a summary of a foreign data base, the Swedish T-book. Section 4.2.5 contains a discussion of several non-nuclear data bases which could be useful for some data issues in nuclear power PRA. Section 4.2.6 describes a process for selecting a generic data value from these sources.

4.2.1 NRC-Sponsored Generic Data Bases

The discussion of NRC-sponsored generic data bases is presented in two sections. The first discusses current data bases. These data sources are deemed appropriate for current and future use. The second section briefly summarizes some historical data bases that have been used or referenced in past analyses. **While useful at the time, these data bases are no longer considered appropriate sources of information.**

4.2.1.1 Current Data Bases

Current NRC-sponsored data bases are discussed in the following subsections. Major attributes for each data base are identified, and limitations associated with each data base are provided.

As a reminder, these data bases are considered to be appropriate sources of information for use in PRAs or other risk assessments. However, it is the user's responsibility to ensure that any information from these data bases used in their analysis is appropriate for their analysis.

4.2.1.1.1 Severe Accident Risks Study Generic Data Base (NUREG-1150)

The generic data base developed for the NRC's Severe Accident Risks study (NUREG-1150) (NRC 1990) is documented in NUREG/CR-4550 as supporting documentation (Drouin et al. 1990). This data base was developed from a broad base of information, including:

- WASH 1400 (NRC 1975b),
- the IREP data base (Carlson et al. 1983),
- Zion (ComEd 1981), Limerick (PECO 1982), Big Rock Point (CPC 1981), and the Reactor Safety Study Methodology Application Program (RSSMAP) PRAs (Hatch et al. 1981),
- NRC LER summaries (Hubble and Miller 1980, Appendices O through Y), and
- the NRC's Station Blackout Accident Analysis (Kolaczkowski and Payne 1983).

Component failure probabilities, failure rates, and initiating event frequencies typically modeled in the NUREG-1150 plant analyses are included in the data base. A mean value and an error factor on a log normal distribution are provided for each entry into the data base.

Limitations in the Data Available from NUREG-1150

The basis of the NUREG-1150 data base is from a broad group of prior PRA analyses and generic data bases. Thus, it does not directly represent the results of the analysis of actual operational data. Furthermore, the data upon which those previous analyses are based suffer from limitations similar to those for older NRC data sources and the NPRDS data base (Sections 4.2.1.2 and 4.2.3.1).

4.2.1.1.2 Evaluation of Loss of Offsite Power Events at Nuclear Power Plants: 1980 - 1996

The report *Evaluation of Loss of Offsite Power Events at Nuclear Power Plants: 1980 - 1996*, NUREG/CR-5496 (Atwood et al. 1998), presents an analysis of loss of offsite power (LOSP) initiating event frequency and recovery times for power and shutdown operations at commercial nuclear power plants. The evaluation is based on LERs for events that occurred during 1980 through 1996. The primary objective of the study was to provide mean and uncertainty information for LOSP initiating event frequencies and recovery times. A secondary objective was to re-examine engineering insights from NUREG-1032 (a LOSP study covering the years 1968 through 1985) using the more recent data.

The major findings of the report are:

- Not all LOSP events that occur at power result in a plant trip.
- Plant-centered events clearly dominate the LOSP frequency during both power and non-power operational modes.
- Plant-centered LOSP frequency is significantly higher during shutdown modes than during power operation.
- No statistically significant variation among units was found for plant-centered sustained initiating events.
- During shutdown, statistically significant variation among plants was found for plant-centered sustained initiating events.
- Equipment faults were the main contributor (58%) to plant-centered LOSP initiating events that occurred during power operations. Human error accounted for a smaller contribution (23%).
- During shutdown conditions, human error was the dominant contributor (58%).
- A clear downward trend can be seen for the plant-centered initiating event frequency.
- Grid-related LOSP frequency is small.
- For severe weather, statistically significant site-to-site variability exists for sustained shutdown LOSP frequencies.
- Severe weather events had significantly longer sustained recovery times.
- For sustained recovery times, no pattern was found correlating unit design class with longer recovery times.
- Longer recovery times were observed for sustained plant-centered LOSP events that did not result in a plant trip or that occurred during shutdown.

Nominal frequencies and upper and lower bounds are given in the report.

Limitations in the Data Available from NUREG/CR-5496

The generic data base developed in this NRC-sponsored data study is based on raw data from LERs. LERs constitute data involving only reportable events at nuclear power plants, and the degree of detail provided in the LERs varies. Some information needed in the data analysis had to be estimated (e.g., allocation of 1980 time into critical and shutdown time), and the analysis ended with events that occurred in 1996. Thus, the data base does not contain events that occurred after 1996, and may not be representative of actual current operational experience.

4.2.1.1.3 Rates of Initiating Events at U.S. Nuclear Power Plants: 1987 - 1995

The report *Rates of Initiating Events at U.S. Nuclear Power Plants: 1987 - 1995,* NUREG/CR-5750 (Poloski et al. 1999a), presents an analysis of initiating event frequencies at domestic nuclear power plants. The evaluation is based primarily on the operational experience from 1987 through 1995 as reported in LERs. The objectives of the study were to:

- provide revised frequencies for initiation events in domestic nuclear plants,
- compare these estimates to estimates used in PRAs and Individual Plant Evaluations (IPEs), and
- determine trends and patterns of plant performance.

Major findings of the report are:

- Combined initiating event frequencies for all initiators from 1987 through 1995 are lower than the frequencies used in NUREG-1150 (NRC 1990) and industry IPEs by a factor of five and four, respectively.
- General transients constitute 77% of all initiating events, while events that pose a more severe challenge to mitigation systems constitute 23%.
- Over the time period of the study, either a decreasing or constant time trend was observed for all categories of events.
- Loss of coolant accident (LOCA) frequencies are lower than those used in NUREG-1150 and industry IPEs.

Nominal frequencies and upper and lower bounds are given in the report.

Limitations in the Data Available from NUREG/CR-5750

The generic data base developed in this NRC-sponsored data study is primarily based on raw LER data from 1987 through 1995. For some events (e.g., LOCAs) information from additional operating experience, both domestic and foreign, was used with other sources of information (e.g., engineering analyses) to estimate the initiating event frequencies. Since the analysis ended with events that occurred in 1995 and made use of other sources of information, the data base may not be representative of actual current operational experience.

4.2.1.1.4 System Reliability Studies

A series of system reliability studies, documented in the multi-volume NUREG/CR-5500 report,[1] presents an analysis of system unreliability for various systems.[2] The following volumes comprise the systems that will be studied::

- Volume 1: auxiliary/emergency feedwater system (Poloski et al. 1998),
- Volume 2: Westinghouse reactor protection system (Eide et al. 1999a),
- Volume 3: General Electric reactor protection system (Eide et al. 1999b),
- Volume 4: high-pressure coolant injection system (Grant et al. 1999a),
- Volume 5: emergency diesel generator power system (Grant et al. 1999b),
- Volume 6: isolation condenser system (Grant et al. 1999c),
- Volume 7: reactor core isolation cooling system (Poloski et al. 1999b),
- Volume 8: high-pressure core spray system (Poloski et al. 1999c),
- Volume 9: high pressure safety injection system (Poloski et al. 2000),
- Volume 10: CE reactor protection system (Wierman et al. 2002a), and
- Volume 11: B&W reactor protection system (Wierman et al. 2002b).

With the exception of the reactor protection system volumes, the analyses of the other systems are based on information obtained from LERs. For the reactor protection system volumes, the analyses are based on information obtained from NPRDS and LERs.

The analyses: (1) estimate the system unreliability based on operating experience, (2) compare the estimates with estimates using data from PRAs and IPEs, (3) determine trends and patterns in the data, and (4) provide insights into the failures and failure mechanisms associated with the system.

---

[1] Currently, it is expected that some of these reports will be updated with new information.

[2] Train, subsystem or system data can be combined with basic event failure data to obtain improved estimates of component failure rates. A Bayesian method for doing this is described in Martz and Almond 1997.

Unreliability estimates (means and distributions) are provided for the entire system for each plant. In addition, unreliability estimates for major train segments failure modes (e.g., failure to start – pump, driver, valves, and associated piping) are provided. Common cause failure estimates are also provided.

Limitations in the Data Available from NUREG/CR-5500

The information available from this NRC-sponsored data study is based on that available from LERs and NPRDS. LERs constitute data only involving reportable events at nuclear power plants, and the degree of detail provided in the LERs varies. The limitations associated with NPRDS are provided in Section 4.2.3.1. The information used in the studies spans various time frames, with the most up-to-date information coming from 1997. Thus, the results of the studies may not be representative of actual current operational experience.

4.2.1.1.5 Component Performance Studies

A series of component performance studies, documented in the multi-volume NUREG-1715 report, presents an analysis of component performance for various components. The following volumes comprise the components that have been studied:

- Volume 1: turbine-driven pumps (Houghton and Hamzehee 2000a),
- Volume 2: motor-driven pumps (Houghton and Hamzehee 2000b),
- Volume 3: air-operated valves (Houghton 2001a), and
- Volume 4: motor-operated valves (Houghton 2001b).

The analyses are based on information obtained from NPRDS and LERs. The data included in the studies cover the period 1987 through 1995.

The analyses: (1) estimate the system-dependent unreliability of selected components, (2) compare the estimates with estimates from PRAs and IPEs, (3) determine trends and patterns in the data, and (4) provide insights into component performance, including component failure mechanisms.

System-dependent unreliability estimates (means and distributions) for various failure mechanisms are provided for each component. Trends in component failure rates were also evaluated in these studies.

Limitations in the Data Available from NUREG-1715

The information available from this NRC-sponsored data study is based on that available from LERs and NPRDS. LERs constitute data only involving reportable events at nuclear power plants, and the degree of detail provided in the LERs varies. The limitations associated with NPRDS are provided in Section 4.2.3.1. The information used in the studies spans various time frames, with the most up-to-date information coming from 1998. Thus, the results of the studies may not be representative of actual current operational experience.

4.2.1.2 Historical Data Bases

In the past, NRC sponsored several programs to develop data bases on nuclear power plant component reliability and initiating event frequencies. These programs included:

- In-Plant Reliability Data Base for Nuclear Power Plant Components (IPRDS) (Drago et al. 1982) – established at Oak Ridge National Laboratory to establish methods for data collection and analysis.
- Nuclear Reliability Evaluation Program (NREP) – generic data base developed to support the Probabilistic Safety Analysis Procedures Guide, NUREG/CR-2815 (Papazoglou et al. 1984).
- Interim Reliability Evaluation Program (IREP) Generic Data Base – developed to support the performance of five PRAs in the 1980s and documented in the IREP procedures guide (Carlson et al. 1983).
- Nuclear Computerized Library for Assessing Reactor Reliability (NUCLARR) – developed as a repository of human error and hardware failure information that could be used to support a variety of analytical techniques for assessing risk. NUCLARR was documented in five volumes as NUREG/CR-4639 (Gertman et al. 1990).

Major attributes for each program and the resulting data bases are documented in the cited references.

4.2.2 DOE-Sponsored Generic Data Bases

Several data bases have been developed to support DOE-sponsored projects. Two of these data bases are discussed in the following sections.

4.2.2.1 Component External Leakage and Rupture Frequency Estimates

Estimates of external leakage and rupture frequencies for components such as piping, valves, pumps, and flanges are necessary for detailed risk analysis of internal flooding. These estimates have been developed and documented in EGG-SSRE-9639 (Eide et al. 1991). The estimates are based on an analysis of data gathered from a comprehensive search of LERs contained in Nuclear Power Experience (NPE) (Hagler-Bailly 1972).

The NPE data base was searched for data covering the period September 1960 through June 1990. The external leakage and rupture events collected from the data were converted to component leakage and rupture frequencies in a three-step process:

1. The ratios of external rupture events to external leakage and rupture events were examined for various components by size and system to decide how to group the data.
2. The final probabilities of an external rupture, given an external leakage or rupture event, were determined.
3. Lastly, the external leakage and rupture frequencies were obtained by estimating component populations and exposure times.

Limitations in the Data Available from EGG-SSRE-9639

The generic data base developed in this DOE-sponsored data study is based on raw LER data from 1960 through 1990. LERs constitute data only involving reportable events at nuclear power plants, and the degree of detail provided in the LERs varies. Since the analysis ended with events that occurred in 1990, the data base may not be representative of actual current operational experience.

4.2.2.2 Generic Component Failure Data Base for Light Water and Liquid Sodium Reactor PRAs

A generic component failure data base was developed by the Idaho National Engineering Laboratory (INEL) for light water and liquid sodium reactor PRAs. This data base is documented in EGG-SSRE-8875 (Eide et al. 1990). The intent of this project was to base the component failure rates on available plant data as much as possible rather than on estimates or data from other types of facilities. The NUCLARR data base and the Centralized Reliability Data Organization (CREDO) (Manning et al. 1986) were used as the primary sources of component failure data. If specific components and failure modes were not covered in those two sources, then other standard sources such as IEEE STD-500 (IEEE 1983) (for electrical components) and WASH-1400 (NRC 1975b) were used. The data base is organized into four categories according to the working fluid of the component:

- mechanical components (water or steam),
- mechanical components (liquid sodium),
- mechanical components (air or gas), and
- electrical components.

Limitations in the Data Available from EGG-SSRE-8875

The generic data base developed in this DOE-sponsored data study is based on information from multiple sources. Since the analysis ended with events that occurred in 1990, the data base may not be representative of actual current operational experience.

4.2.3 Industry Data Bases

Several data bases developed within the nuclear power industry for both risk assessment and for plant operations are summarized here. Data bases discussed in this section were developed by the Electric Power Research Institute (EPRI) and the consulting firms of EQE, International and Science Applications International Corporation (SAIC).

Although the NPRDS and EPIX data bases (described in Section 4.1.3) contain plant-specific data, they can be used to generate generic failure rates for components. Methods for aggregating individual plant data to estimate failure rates are described in Section 8.2 of this handbook. Aggregation of data from EPIX can be performed using the RADS software developed under the NRC auspice.

4.2.3.1 EQE, International

The EQE, International generic data base (formerly known as the Pickard, Lowe, and Garrick or PLG data base) for light water reactors is set up to support PRA and reliability analysis for which both point estimates

and uncertainty distributions are developed.[3] The data base contains information on:

- Component failure rates,
- Common cause failures,
- Component maintenance frequencies and mean durations,
- Initiating events,
- Fire and flood events at nuclear sites, and
- Shutdown events involving loss of residual heat removal (RHR) cooling and loss of inventory.

The fire, flood, and shutdown events are a compendium of experience event summaries from all U.S. nuclear sites. The common cause data are presented as event description and have been classified according to the methodology of NUREG/CR-4780 (Mosleh et al. 1989). The fire, flood, shutdown and common cause events have, in addition to the description, information in various fields making them convenient for sorting and for use in plant-specific screening analysis.

All other data are in the form of distributions and are compatible with the PLG risk assessment software, RISKMAN® . These distributions are generated using the data analysis module of RISKMAN® which can be used as a stand-alone software. The distributions developed are available to the other modules of RISKMAN® used for fault-tree quantification and core damage sequence quantification.

The actuarial data are from over 20 nuclear sites in the U.S. and in Europe. Other sources of generic information also used are:

- EPRI reports on components, shutdown accident events, initiating events, loss of offsite power;
- Special NUREG reports on components such as pumps, valves, diesel/generators;
- Compiled data bases such as Nuclear Power Experience, NUCLARR, IEEE-500 (IEEE 1983), NPRDS, etc.; and
- Insurance company databases for loss events.

The database includes statistics for components that cover population, demands, operating times, failures, and maintenance outages and durations at specific plants. It also includes event-by-event analyses for initiating events, common cause failures, and fires and floods over the whole U.S. plant population. In addition to this factual information, parameter estimates from published sources of generic reliability data are also provided.

The actuarial data and the other generic data are combined using a two-stage Bayesian updating technique. The generic distributions maintain what is referred to as plant-to-plant variability. Since the data are developed specifically to be used for Monte Carlo sampling, they are defined with a minimum of 20 discrete bins with special attention given to the tails of the distributions.

The database is available in a format compatible with RISKMAN® and also as ASCII files.

Limitations in the Data Available from EQE, International

The EQE data base is proprietary, so the adequacy and comprehensiveness of the underlying data have not been evaluated for this document. As noted above, several of the sources of generic information incorporated into the data base are discussed previously in this chapter (e.g., NUCLARR, NPRDS); thus, it is possible that some of the data from the EQE data base may have limitations similar to other data bases discussed in this chapter. However, it should be noted that the proprietary nature of the EQE data base precludes any definitive judgment as to how data bases such as NUCLARR and NPRDS were utilized in the development of the EQE database.

4.2.3.2 Science Applications International Corporation

Science Applications International Corporation (SAIC) has developed a generic, proprietary data base for application to PRAs on commercial nuclear power plants.[4]

The scope of the data base for components and their failure modes was established by a review and tabulation of all basic events and component failures in SAIC-conducted PRAs. Components were grouped

[3] The information on the EQE/PLG data base is based on personal correspondence from Shabha Rao, PLG, Newport Beach, California, to Timothy Wheeler, Sandia National Laboratories, September 16, 1999, and to Donnie Whitehead, Sandia National Laboratories, April 4, 2001.

[4] The information on the SAIC data base is based on a personal correspondence from Alan Kolaczkowski, Vice President, SAIC, to Donnie Whitehead, Sandia National Laboratories, April 18, 2001.

into generic categories rather than specifically by system or application. Thus, all basic events for motor-driven pumps were categorized into a single "motor-driven-pump" category rather than delineated size or by system. Some component failure modes were merged to reflect the available data (e.g., air-operated valves fail-to-open and fail-to-close were combined into a single failure mode – fail-to-operate. Component boundary definitions are given for all components in the SAIC generic data base.

The data base was developed by collecting all sources of available parameter estimates relevant to the component failures defined by the scoping process. Each data source was evaluated against a set of acceptance criteria, including availability (no proprietary sources were included), compatibility of data to being fit to a lognormal distribution, and Bayesian updating. Any source which used Bayesian parameter estimation methods to develop estimates for component failure modes was rejected. Such data sources were considered to be too plant-specific for inclusion into a generic data base.

Each individual data source selected against the acceptance criteria was fitted to a lognormal distribution. Then, all data sources for each particular component failure were aggregated through a weighted sum approach (each source was weighted equally). Each aggregated distribution was fitted to a lognormal distribution.

Limitations in the Data Available from the SAIC Data Base

The SAIC data base is proprietary, so the adequacy and comprehensiveness of the underlying data have not been evaluated for this document.

4.2.3.3 Advanced Light Water Reactor Data Base

EPRI's Advanced Light Water Reactor (ALWR) Utility Requirements Document (EPRI 1989) contains a reliability data base for use in ALWR PRAs. Several data sources were reviewed and representative failure rates and event probabilities were compiled from these data sources. A best estimate value was selected for each component type and failure mode based on judgment regarding the applicability of the data source to the expected ALWR design. The primary sources used in the data survey were the Oconee PRA (Duke 1984), the Seabrook Probabilistic Safety Study (PLG 1983), parameter estimates from licensee-event reports documented in NUREG/CR-1363 (Battle 1983) for valves, NUREG/CR-1205 (Trojovsky 1982) for pumps, and NUREG/CR-1362 for diesel generators (Poloski and Sullivan 1980).

Limitations in the Data Available from the ALWR Data Base

The ALWR data base lists only best estimates for each initiating event, failure rate, and event probability. The survey is well documented in that all estimates collected for each parameter estimate are shown. However, only a cursory statement of rationale for deriving the best estimate value is given. No uncertainty bounds or probability density functions are given.

4.2.4 Foreign Sources

Two sources of data from Nordic nuclear power plants are available. The I-Book documents initiating event frequency data and the T-Book documents component failure data.

4.2.4.1 Sweden's T-Book for Nordic Nuclear Power Plants

Since the early 1980s a Reliability Data Handbook, the T-Book (ATV 1992), has been developed and used for nuclear power plant of Swedish design. The T-Book provides failure data for the calculation of component reliability for use in regulatory safety analyses of Nordic nuclear power plants. The 3rd edition is based on operation statistics from 12 Swedish and 2 Finnish nuclear power plants, including approximately 110 reactor years of experience.

The failure characteristics incorporated into the parameter estimations in the T-Book are based on Licensee Event Reports for Nordic plants delivered to the Swedish Nuclear Power Inspectorate (SKi) and from failure reports in ATV's central data base. Only critical failures, those that actually caused a component's function to stop or fail, are incorporated into the parameter estimations. A multistage empirical Bayesian approach is used to develop the component parameter estimates from the raw data (Pörn 1996).

Limitations in the Data Available from the T-Book

Data for the T-Book are collected from LERs delivered to the SKi; thus, the parameter estimates derived from the data are based only on data of reportable incidents. It is not understood how representative such data may be of actual operational experience.

4.2.4.2 Sweden's I-Book for Nordic Nuclear Power Plants

The I-Book (Pørn et al. 1994) contains a compilation of initiating events that have occurred in Nordic nuclear power plants. The data reflects 215 reactor years of operating experience prior to 1994. In the first edition of the I-Book, issued in 1993 (Pørn et al. 1993), initiating event groups were identified and frequencies generated. The operating experience from two additional plants in Finland were included in the second edition (Pørn et al. 1994).

The I-Book includes the development of a statistical model for performing a trend analysis. The model is based on nonhomogeneous Poisson (Power Law) processes and includes a complete treatment of parametric uncertainty using Bayesian methods.

Limitations in the Data Available from the I-Book

Data for the I-Book are collected from operating experience at Nordic plants. It is not understood how representative such data may be of operational experience in nuclear power plants in the United States.

4.2.5 Non-Nuclear Power Data Bases

There are many non-nuclear data bases that contain failure data that can potentially be used in nuclear power plant PRAs. Several of these data bases are described below. When using data from non-commercial nuclear sources, care must be taken to ensure that the data are for components and conditions representative of those that exist in nuclear power plants.

4.2.5.1 Reliability Analysis Center

The Reliability Analysis Center (RAC) in Rome, New York, maintains two data bases on electronic and non-electronic component reliability. The data bases are:

- Electronic Parts Reliability Data (Denson et al. 1996), and
- Non-Electronic Parts Reliability Data (Denson et al. 1995).

These RAC databases provide empirical field failure rate data on a wide range of electronic components and electrical, mechanical, and electro-mechanical parts and assemblies. The failure rate data contained in these documents represent cumulative compilation from the early 1970s up to the publication year for each document. Data are collected from sources such as:

- published reports and papers,
- government-sponsored studies,
- military maintenance data collection systems,
- commercial/industrial maintenance databases, and
- direct submittals to the RAC from military or commercial organizations that maintain failure data bases.

Limitations in the Data Available from the RAC Handbooks

The RAC handbooks provide point estimate parameter estimations for failure rates (or demand probabilities). No treatment of uncertainty is provided.

4.2.5.2 Offshore Reliability Data Project

The Offshore Reliability Data (OREDA) project has collected and processed data from offshore oil platforms operated by 10 different companies off the coasts of the U.K., Italy, and Norway. Reliability data collected and processed by OREDA has been published in the Offshore Reliability Data Handbook (OREDA 1997). The main objective of OREDA is to collect reliability data for safety important equipment in the offshore oil industry.

Components and systems for which data are collected are:

- Machinery
  - Compressors
  - Gas turbines
  - Pumps
- Electric generators
- Mechanical Equipment
  - Heat exchangers
  - Vessels
- Control and Safety Equipment
  - Control Logic Units
  - Fire and Gas Detectors
  - Process sensors
- Valves
- Subsea Equipment
  - Control Systems
  - Well completions

Data have been collected from 7,629 individual equipment units (e.g., individual pumps, valves, motors) over a total observation period of 22,373 years. The records include 11,154 failures.

Under each category of equipment (e.g., machinery) information is collected on each type of component (e.g., centrifugal compressors). Data are further sorted by a component's driving mechanism (e.g., electric motor-driven), by failure mode (e.g., fails-to-start, fails-while-running), and by the criticality of each failure (e.g., critical - terminates the operation of the component, degraded - component still operates).

The OREDA-97 handbook presents failure rate and demand failure probability estimates for various combinations of component function, application, capacity, operating fluid, and size.

Limitations in the Data Available from the OREDA Data Base

Certain data quality issues have arisen in the development of OREDA (Sandtorv et al. 1996). The quality and availability of data can vary significantly among the 10 participating companies. Interpretations of equipment definitions and failure mode specifications can vary among the participants as well, affecting the quality of data. The effect of preventive maintenance on equipment reliability is difficult to measure. Since preventive maintenance practices vary among the participating companies it is unclear as to what would be the baseline rate of a generic type of equipment.

4.2.5.3 IEEE-500 Standard

The Institute of Electrical and Electronics Engineers (IEEE), Inc. Standard 500-1984 (IEEE 1983) contains failure estimates for various electrical, electronic, sensing, and mechanical components. Delphi procedures (an elicitation process) were used in producing component failure estimates. Multiple sources of information, including nuclear, fossil fuel, and industrial, were considered by the experts as part of the Delphi process.

Limitations in the IEEE-500 Data Base

The major limitations associated with the IEEE-500 data base are (1) the data base contains dated material (i.e., the latest information used to develop the data base comes from the early 1980s), and (2) the process used to support development of the failure estimates was an uncontrolled process. (A survey was sent to various individuals requesting them to provide information on selected issues. No inherent controls were placed on the individuals, and no training on how to estimate failure probabilities was provided to the individuals filling out the survey forms.) In addition, it should be noted that IEEE Standard 500-1984 has been withdrawn and is no longer available from IEEE.

4.2.6 Selection of Parameter Estimates from Existing Data Bases

The need to select parameter estimates from existing generic data bases may arise when performing a PRA. This can occur when a PRA is being performed on a new plant that has no operating history or it may occur when no plant-specific information exists for a specific component. Whatever the reason, when it becomes necessary to select parameter estimates from generic data bases, certain cautions should be observed:

1. The generic data base should contain failure probability estimates for components that are identical or comparable to the ones in the PRA model in terms of size, component boundary definition, intended operational history (e.g., normally operating versus standby), and expected or postulated operating environment.

2. The generic data base should contain a recommended point estimate and an uncertainty distribution for each identified failure.

3. If possible, the primary sources of information used to develop the generic data base's failure probabilities and distributions should be information from other nuclear power plants. Supplemental information from non-nuclear sources should be used only when necessary to provide failure probabilities and distributions for components that cannot be obtained from nuclear power plant generic data sources.

4. Where possible, the generic data base's failure probabilities and distributions should be derived from actual failure events. If such information is not available, then failure probabilities and distributions generated by other techniques (e.g., expert elicitation) are acceptable.

5. Generic data base failure probabilities and distributions should reflect current trends. If significant trends exist within the failure data indicating either an increase or decrease in the failure probabilities, the underlying event failure information used to generate the failure probabilities should represent these recent events.

However, if no significant trends exist, then data from all years can be used to estimate the failure probabilities.

6. The failure probability estimates contained within the generic data base should not be based on incestuous sources, i.e., the estimates should not be derived from two different sources that employed similar or different analysis techniques to the same ultimate set of failure information.

5. PLANT-SPECIFIC DATA COLLECTION AND INTERPRETATION

The incorporation of plant-specific data in the parameter estimates used in a PRA produces risk estimates that reflect the actual plant experience. A plant-specific data analysis also allows comparison of plant equipment performance relative to an industry average (the generic value). A plant-specific data analysis will identify those components or systems whose performance is worse than the industry average. It may also identify components or systems with better-than-average performance.

As indicated in Chapter 4, the raw failure data needed for a plant-specific data analysis is dependent upon the scope the analysis. The scope can include accident initiating events, component failure events and unavailablilities due to maintenance or testing, and recovery events. Typical sources of raw data available at nuclear power plants for each of these type of events are identified in Section 4.1. The information needed may have to come from multiple sources.

Interpretation and reduction of the raw data is required to obtain the reduced data used in the parameter estimation models described in Chapters 2 and 6. The reduction of the raw data includes consideration of issues such as pooling of identical component data, the mode of operation the plant was in when a failure occurred, and the severity of the event. Additional issues concerning data reduction, such as aging and time impacts, are addressed in Chapter 7.

This chapter describes a process for collecting and reducing raw data for the purpose of generating plant-specific data for use in a PRA. Because nuclear power plants collect and record raw data in different ways, the process described is general in nature but, sufficient to successfully collect and reduce available date for use in a PRA. Some practical concerns and issues related to the scope and performance of plant-specific data analysis are also presented.

A process for reducing the data necessary to calculate initiating event frequencies, component failure data, and recovery event data are presented in Sections 5.1, 5.2, and 5.3, respectively. The reduced data obtained in this process are combined according to the guidance provided in Chapters 2 and 6 to obtain the parameters necessary to quantify PRA models.

5.1 Initiating Event Data

The methods for evaluating plant-specific initiating event frequencies provided in Chapter 6 require the number of initiating events of interest and the time period over which these events occurred. Guidance is provided in this section for collecting and interpreting this required data.

5.1.1 Initiating Event Categories

The initiating events of interest in nuclear power plant PRAs are dependent upon the mode of operation that the plant is in. For power operation, the events of interest are generally reactor scrams but can also include unplanned forced shutdowns. Typical initiating events during power operation include multiple categories of plant transients and loss-of-coolant accidents (LOCAs). Trips from zero power or low power may be excluded as valid initiating events in a full power PRA if their occurrence is precluded during full power operation. However, low power events should be considered as valid initiating events at full power if they can occur during full power. For shutdown modes of operation, the reactor is already subcritical and thus the events of interest are somewhat different. Typical initiating events modeled in shutdown PRAs include loss of decay heat removal events, reactivity insertion events, and LOCAs or drain-down events.

It is a standard practice in PRAs to group initiating events into categories based on their impact on certain plant systems, and according to the demands they make on other plant systems needed for accident mitigation. Examples of typical initiating event categories include loss of offsite power, loss of feedwater, main steam isolation valve (MSIV) closure, and large, medium, and small LOCAs. Lists of typical transients that have occurred at nuclear power plants while at full power have been categorized by EPRI (1982) and the INEEL (Mackowiak et al. 1985 and Poloski et al. 1999a). Typical initiating events to consider during low power and shutdown conditions have also been established for both boiling water reactors (BWRs) (Staple et al. 1999) and pressurized water reactors (PWRs) (Chu et al. 1993).

5.1.2 Data Window

The time period for collecting initiating event data should be as broad as possible. In general, data from all of the years of plant operation should be considered. However, screening of the data can be performed to eliminate unrepresentative events (see the next section). One screening technique used in general practice is to eliminate the first year of operational data as unrepresentative.

Since the number of plant events can decrease over time due to improvements in the design and operation of the plant, it is desirable to have the data reflect the most recent operating experience. This can be accomplished by considering only the data from the most recent years of operation. Alternatively, an analyst could perform a trend analysis of the data (see Chapter 7).

5.1.3 Initiating Event Data Allocation and Screening

To allocate plant-specific event data to the initiating event categories modeled in the plant PRA, it is necessary to establish the status of the plant, including its power level at the time of the event and the impact of the event on the plant systems. Such information is generally available in the raw data sources discussed in Section 4.1 that are available to identify initiating events (i.e., LERs, scram reports, and monthly operating reports).

For initiating events during power operation, the events of concern are those that result in a reactor trip or forced shutdown. To allocate these events to the appropriate initiating event category, a data analyst must examine the sequence of events prior to and immediately following the reactor trip/shutdown. The initial plant fault leading to a sequence of events that eventually result in an automatic or manul reactor trip or unplanned shutdown is used in categorizing the event. For example, one plant trip may have been initiated by spurious closure of the MSIVs and be identified as an MSIV closure transient. Another event may be initiated by a loss of condenser vacuum which produces a closure of the MSIVs. This event may also be placed in the MSIV closure transient category, unless some significant difference in the plant response is identified.

The initiating event data analysis can also be used to help establish the conditional probability of events subsequent to the event actually leading to the plant trip. Examples of this include the failure of the reactor protection system leading to an anticipated transient without scram (ATWS), and the occurrence of a relief valve sticking open leading to a transient-induced LOCA.

It is possible that some events leading to plant scrams (or loss of heat removal during a shutdown mode of operation) can be eliminated from the data analysis. One acceptable reason for eliminating initiating event data involves design or operational changes that may have been made to reduce the frequency of reactor scrams. Such changes to the plant design or operation can eliminate the occurrence of failures that have occurred in the past. For example, a plant may have experienced a significant number of loss of feedwater events due to the design of the feedwater control system. As a result, a utility may have replaced the feedwater controller with a new, more reliable design that eliminated the occurrence of loss of feedwater due to controller faults. The data analyst can thus eliminate past events initiated by faults in the old feedwater controller from consideration.

Changes in the plant design or operation can also affect the classification of events. The following example, provided in EPRI TR-100381 (EPRI 1992), illustrates this point. The MSIV vessel level closure set point at some BWRs has been lowered from Level 2 to Level 1. As a result, the fraction of initiating events that lead to MSIV closure may be different before and after the design change implementation and the total historical count of MSIV closure events may not be valid for the current condition of the plant. One approach for dealing with such a design change is to eliminate all events prior to the design change that result in MSIV closure due to the generation of a low vessel level. This approach has the undesirable impact of reducing the sample size. An alternative is to review the past events to determine if the MSIVs would have closed with the revised closure set point in place. However, this may be difficult to determine from the available information.

5.1.4 Selection of Exposure Time

For estimating the frequencies of initiating events that occur during any plant operating mode, the appropriate exposure time is the number of calendar years of operation corresponding to the period of time the initiating event data is collected. Expressing the frequency of initiating events on a calender year basis allows for evaluation of risk in each mode on a consistent and average basis.

However, it may be necessary to generate the initiating event frequencies based on the time the plant is in the particular mode of operation. For example, initiating events during power operation are often expressed in terms of events per critical year (one critical year represents 8760 hours of reactor criticality). Since generic initiating event frequencies are often expressed in events per critical year (Poloski 1999a), calculation of the plant-specific frequencies in this same unit is required for combining the two values using Bayesian techniques (see Section 6.2.2). To determine at-power initiating event frequencies, the plant-specific frequencies expressed as events per calender year have

to be increased by dividing by the fraction of time the plant was at power. This fraction is called the **criticality factor** and may be determined from the control room logs or the Grey Books where the residence times in each of the operational modes are recorded. Criticality factors for each plant are provided in Appendix H of NUREG/CR-5750 (Poloski 1999a) for the years 1987 through 1995. Alternatively, the generic frequencies may be divided by the average criticality factor (0.75 for the data reported in NUREG/CR-5750) to obtain generic data expressed in the same units as the plant-specific data (i.e., events per calender year. For example, suppose an event is expected to occur 1.6 times every calender year, on average, and that the criticality factor for a specific plant is 0.8 (i.e., the reactor has been critical 80% of the time). Then, the same event correlated to units of critical years is 2 events per critical year (1.6 events/calender year divided by 0.8 critical years/calender year).

5.2 Component Failure Data

The raw data sources containing equipment operating records in a nuclear power plant typically document tens of thousands of component malfunctions over the plant's lifetime. The records may be kept in various forms including hard copies of maintenance work orders or a computerized file. The most useful raw data sources provide information on the specific component affected, the observed problem, and the action taken. To calculate plant-specific component failure rates and unavailability from the data in these records, the data analyst must identify those malfunctions that cause component functional failures and also determine the corresponding number of demands or operating time. This section describes this process and some of the practical concerns required to extract the necessary data.

5.2.1 Component Data Identification

The first step in evaluating plant-specific component failure rates is to identify the components and their failure modes that will be analyzed. This step is usually done in coordination with other PRA analysts (typically those analysts that generate system models such as fault trees). This coordination is critical because it focuses the component data analysis on only those components and their failure modes that appear in the PRA models and establishes the definitions of the component boundaries.

It should be noted that extremely reliable components may never have failed in the history of the plant. This lack of failure history makes it difficult to estimate the true failure rate or probability. Reliable components can generally be identified by reviewing failure rates in generic data bases. However, the analyst is cautioned in the use of this data since a usually reliable component may not be reliable at a particular plant. In addition, it is often impossible to identify the number of demands or run times for certain components (for example, the number of demands placed on a relay) using the existing plant records.

5.2.1.1 Data Window

Plant-specific data is selected over a sufficient time period to provide statistically meaningful results. Use of data from throughout the plant history is preferred since they will be less subject to random variability. The following examples from EPRI TR-100381 (EPRI 1992) illustrates the amount of data required to achieve an acceptable sample size.

> "With no failures, the statistical significance can be measured by the 95th upper confidence limit. To establish a 95th confidence limit on a failure rate of 1E-3/hr, the required cumulative run time for the population is 3,000 hours, to establish a 95th confidence limit of 1E-4/hr requires 30,000 hours. Thus, if a failure rate is believed from generic data to be relatively low, one should expect to have to collect a significant amount of run time before making an impact on the generic values.
>
> "When failures are recorded the statistical significance can be measured by the range from the 5th to the 95th percentile confidence bounds. This decreases with the number of failures. For a Poisson distribution, the range from the 5th to the 95th percentile is on the order of 10, with 2 failures. Thus, for greater than 2 failures the sample is very loosely comparable to the lognormal with an error factor of 3. Thus, for a population of components, a total number of failures of 2 or more is a reasonable sample when compared with typical generic data bases. This is true for the binomial distribution also, as it approximates the Poisson distribution when the parameter, p, is on the order of 10^{-3}. These considerations can be used to establish a reasonable time frame for data collection. Suppose, the generic data is on the order of $l0^{-3}$per demand, and there are four components in

the population with approximately one demand per component per month per ISI tests. To get 2 failures, we would expect to require about $2/p$ demands, or 2,000 demands. There are 48 demands per year, therefore data from 41 years would be required to produce this statistically meaningful data. This illustrates the importance of making sure that all the demands are counted and also of increasing the size of the population if at all possible."

5.2.1.2 Data Collection

For the list of components and their failure modes selected for data analysis, the system analyst must retrieve all failure, maintenance, and test records for each component from the raw data sources generated during the data window. The required records are generally obtained based on the component identification number. Because the component boundary can include multiple piece parts, the required records may be kept under multiple identification numbers. However, for some components, the data records for the different piece parts may all be kept under the same identification number. Thus, it is necessary to list the identification numbers for all the piece parts included in the component boundary definition.

Because component failures are generally infrequent, it is preferable to pool the data from several components to obtain a larger data base. For example, it is common to group like pumps within a single system into one population, but less common to group the pumps of different systems (although it can be acceptable to group pumps of different systems with similar characteristics together into one population). Any grouping of components requires careful consideration of the similarity of their design (e.g., size or manufacturer), the frequency of operation, their environmental operating conditions (e.g., temperature, humidity, and radiation), operating modes (e.g., standby versus normally operating or intermittently operating), and the medium they carry (e.g., air, pure water, or borated water). Tests for poolability of data are described in Chapter 6.

5.2.2 Event Screening and Severity Classification

The raw data for a specific component will contain some events that are not relevant to the component failure modes being analyzed. These events can be screened from further analysis. Some of the events will be component failures that should be included in the data assessment. The type of component failures will determine how they are classified and subsequently used to generate the required component failure data. Guidance for both event screening and classification is provided below.

5.2.2.1 Event Screening

One consideration in the identification of plant-specific data is whether design changes have been made to the plant or its components that invalidate some of the historical data. For example, changing the type of flow controller could impact the operation of a particular turbine-driven pump. Thus, the total historical count of the turbine-driven pump events is not valid for the current condition of the plant. Typically, the turbine-driven pump data prior to the design change would be deleted from the data analysis. However, this has the undesirable impact of reducing sample size. Another approach is to investigate whether there is indeed a significant difference in the fraction of events before and after the design change. Not all the failures may be invalidated by the design change and so the historical data prior to the design change implementation may have partial validity and could be included in the data analysis.

Consideration of design changes is one example of where censoring of data can and should be performed. Other reasons can be used for data censoring if they are well supported and valid. For example, it is not uncommon to eliminate data from the first year of plant operation since it represents failures that occurred during the plant break-in period. However, any data censoring should be approached carefully to avoid losing important information and biasing results (eliminating the first year of data actually makes the results less biased).

5.2.2.2 Event Severity Classification

As discussed in Chapter 3, component malfunction events are commonly classified into one of the following three event severity categories:

- catastrophic failures,
- degraded failures, and
- incipient failures.

Catastrophic failures require some kind of repair or replacement action on the component in order to restore the component to operability. Events that are classified as catastrophic failures are used in calculating plant-specific component failure rates and probabilities of

failure on demand. Information on catastrophic failures occurring during critical operation is also used in calculating maintenance outage unavailabilities.

Degraded failures can prevent a system or train from meeting the success criteria modeled in the PRA. An incipient failure is such that there is no significant degradation in performance but there are indications of a developing fault. The difference between the two is generally a matter of severity. Events classified as incipient or degraded failures are generally used in calculating plant-specific maintenance unavailabilities. Although both degraded and incipient failures will typically lead to a corrective action, the corrective action may or may not make the component unavailable to perform its function. For example, maintenance on the operator of a valve that is normally open will not lead to the unavailability of the valve if is required to open for system operation. This illustrates the importance of ascertaining from event records the modes of a component operation that a corrective action would prevent.

Sometimes the event information is so unclear and incomplete that a definite classification of the severity of a component malfunction event is not possible. The data analyst in this situation is faced with the difficult task of deciding whether to call a malfunction a failure or not. The inability to distinguish between severity levels of failures is particularly important. The difference between the probabilities of catastrophic and degraded modes of failures can be significant especially when dealing with highly reliable components that rarely fail. The difference between no failures and one failure in estimating the failure rate is much more than the difference between 10 and 11 failures. Thus, the data analyst must be careful when classifying the few failures that may have occurred. In the absence of sufficient information, the tendency is to conservatively record such events as catastrophic failures. This is reasonable as long as the impact on the final PRA results is not significant. For cases where the judgement of the data analyst is important to the PRA results, it could be incorporated explicitly into the PRA quantification as a source of uncertainty. This issue is discussed further in Section 6.1.2.2.

5.2.3 Component Data Allocation

This section gives guidelines on the allocation of plant specific events to each component failure mode of interest. This includes the allocation of events contributing to the unavailability of components or systems due to test and maintenance actions. The goal of this allocation process is to correlate each event report with one or more basic events of the PRA model. This requires that the event report be identified with a specific component, and that the severity of the event be determined and associated with the proper component failure mode(s).

The use of component identification numbers in event reports is generally sufficient to allocate the event to a particular component. The description of the event can also guide the data analyst to a particular component failure mode (i.e., a basic event in a fault tree), or in some cases, to a particular gate in a fault tree. However, a thorough review of the cause of the event together with a knowledge of the boundaries of the basic events of the fault trees is generally needed for a correct allocation to be made. For example, an event report identified with a specific motor-operated valve (MOV) that involves the deenergization of a 480V bus should be associated with the bus unavailability and not the MOV. If the event is a local fault of the MOV or its breaker, it is associated with MOV itself.

As discussed previously, the severity of the event is important in allocating the event to specific component failure modes. A catastrophic component failure will generally result in an extended period during which the component is unavailable while it is being repaired. Thus, an event involving a catastrophic failure must be counted in estimating the failure of the component to operate and in estimating its unavailability due to maintenance. Degraded and incipient failures are used in calculating plant-specific maintenance unavailabilities. Some degraded failures may result in sufficient degradation that it can not meet its required success criteria (e.g., the flow rate for a pump is reduced to 300 gpm when 500 gpm is required for success). In such cases, a degraded failure is also included as a component failure to operate.

5.2.3.1 Component Failure Event Allocation

Because of the variability in the level of reporting associated with maintenance events, the allocation of event reports to specific PRA model events can be a subjective process. The following are some ground rules to help perform the component failure event allocation. The majority of these ground rules have been identified and published in EPRI TR-100381 (EPRI 1992). Additional guidelines are based on the experience of PRA vendors and NRC data analysts.

1. For standby components such as pumps, diesel generators, and fans, PRA models generally distinguish between failure to start and failure to run modes. It is important to understand the definition of each failure mode in order to associate historical maintenance events with the different basic event types. For example, if a fault tree basic event represents a failure of a pump to start, it usually means exactly that. However, it is not unusual in PRAs to define "diesel generator fails to start" as encompassing a failure to start or a failure during the first hour given that the start was successful. Whatever definitions are used, the event allocation must be performed to match them.

2. As indicated in Chapter 2, there are two ways to model failures to start: the demand failure and standby failure models. In the demand failure model, the equipment is ready to operate but for some reason, does not start or change state when demanded. In the standby failure model, the equipment has developed an unannounced condition that will prevent it from starting when demanded. When reviewing raw data, it can be difficult to identify whether a component failed on the demand or prior to the demand. Thus, as indicated in Section 2.3.4, either model could be used in this situation. The demand failure model provides the higher failure probability.

3. A catastrophic or degraded failure that is revealed while a component is in the standby mode, and that results in a maintenance action, is accounted for in the unavailability due to maintenance event for that component. If the failure is such that it could also occur while the component is performing its mission, it should also be counted as a component failure. For example, external leakage above allowable amounts from a standby pump that requires isolation of the pump to repair it, contributes to the unavailability of the pump due to maintenance. Since such leakage could occur during pump operation, the event should also be used to determine the failure rate for pump leakage. The amount of leakage would have to be sufficient to prevent the pump train from delivering the required flow.

4. Catastrophic failures of standby equipment to start (or run) that occur during an actual component demand, contribute to that failure mode. Similarly, failures to start (or run) during tests that closely mimics the conditions that the component would be subjected to during an unplanned demand should also be included in the evaluation for the component failure mode.

5. Degraded failures that are not serious enough to prevent the component from performing its function are not included as failures of the component. Expressed in another way, the failure of the component must match the definition of the failure in the PRA model. For example, vibration in a pump that results in the pump only delivering 500 gpm instead of the rated flow of 600 gpm is not a failure event if 500 gpm is sufficient to meet its function and the pump continued to supply that flow for a period at least equal to the mission time required in the PRA model. However, such failures would be included in the unavailability due to maintenance since their effect is to induce maintenance activity.

   There is a caveat to this guideline to consider. If the degraded failure is revealed in a short test duration, an analyst cannot be sure the component would have succeeded over its mission time. In this case, the analyst can attempt to extrapolate the rate of degradation to determine if the component would meet its failure criteria sometime during its mission time. For example: a pump develops a slow oil leak during a test. If the rate of leakage is such that the pump would run out of lubricating oil during the required pump mission time as modeled in the PRA, than the event is considered as a pump failure to continue to run.

6. Degraded conditions for which a failure would have occurred if the system had been demanded are considered a failure. For example, if an operator discovers that a pump had no oil in its lubrication reservoir, the pump may have started (unless there was an interlock preventing a pump start on low oil level) but likely would not have run long. In either case, this event would be counted as a failure to start.

7. If the event report identifies that the failure of component A is the result of the failure of another component B that is modeled explicitly in the PRA, the event is associated with component B and not with component A. For example, failures of a pump from plugged suction screens should not be allocated as pump failures if the screens are modeled separately.

   The clear identification of the component boundary is an important factor in these situations. For

example, the allocation of an event that identifies the failure of an emergency pump due to the failure of a safety actuation signal is dependent upon whether the actuation logic is included in the pump boundary or is treated as a separate event in the model. Typically, the components related to the safety actuation signal are not included in the pump boundary definition and this event should not be counted as a pump failure. However, if the safety actuation signal is included in the pump boundary, then the command fault should be included as a failure mode of the pump.

8. An event reporting a degraded or failed state of a redundant piece part should be excluded from the failure events if the component boundary includes the redundant piece parts. For example, if a diesel generator has two redundant air start motors that are included in the diesel generator boundary definition, failure of one air start motor would not be counted as a failure of the diesel generator. This example illustrates how a coarse definition of a component boundary can result in the failure to account for some degraded component states.

9. If a documented failure during a test or actual demand could not be repeated on subsequent tries, it may not have be included as a potential failure. Similarly, events which are very quickly recoverable may also not be considered potential failures (the recovery should not be included in the PRA model) . Whether an event meeting either of these situations should be considered a failure is a function of the success criterion for the component in terms of the time window within which it has to operate. For example, the spurious closure of an MOV may prevent the injection of coolant into the core from a particular system. However, the event records may indicate that in all such occurrences, the valve was quickly reopened before coolant levels dropped to unacceptable levels. In such cases, the events should not be considered as failure events for the MOV.

10. Successive failures of the same components over short time intervals should be counted as a single failure. Similarly, failures of a component during post-maintenance testing where the failure is related to the maintenance or to an earlier failure that the maintenance was trying to correct should be considered as a continuation of the original failure and should be disregarded. The successive failures are because proper maintenance was not performed to fix the initial problem, and the component is still in the failed state.

11. If failures resulting from human errors after testing, maintenance, and instrument miscalibrations are explicitly included in system models, these events should not be included as component hardware failure events. Such events are typically quantified using human reliability analysis methods. However, some PRAs have not explicitly included these human errors in the models. In such cases, the contribution from human-related failures should be incorporated into the appropriate component failure rate or probability.

12. An event reported as a failure to meet technical specifications, but which would not result in a catastrophic failure in the PRA sense should not be included, but it may lead to a maintenance unavailability. For example, the failure of a diesel generator to start and pick up loads within 10 seconds might be a reportable failure for regulatory purposes. However, in the PRA sense it is not a failure if the diesel did not pick up loads in 10 seconds and the “failure” did not have a discernible effect on the ability of the plant to mitigate an initiating event. However, this failure would require maintenance to alleviate the fast loading failure.

13. Failures that occur under abnormal environmental conditions should be segregated from failures that occur under normal conditions. These failures can identify important interactions between systems and thresholds for failure that should be accounted for in the PRA. In general, PRAs assume components fail under harsh conditions. Under this assumption, actual failure events in harsh environments can be eliminated from consideration. For example, actual failures of electrical components following a loss of a heating, ventilation, or air-conditioning (HVAC) system should be eliminated from the data analysis if the HVAC dependency is modeled explicitly in the PRA model and the component is always assumed to fail under those conditions. However, if there are also many component successes under the same harsh environments, than a component failure probability under those conditions can be calculated and used in the PRA model conditional on the occurrence of the harsh environment.

5.2.3.2 Allocation of Unavailability Data

Unavailability occurs primarily due to maintenance activities but some minor contributions can also result from testing performed during periodic surveillance activities. These unavailability contributions can be

included in a system model at a component, segment, or train level. In addition, separate basic events for maintenance and testing unavailabilities, or for planned and unplanned unavailabilities can be included in system models. In a data analysis, the allocation of unavailability data must be performed to match the basic events in the system models. The following guidelines are useful in allocating events for determining unavailabilities due to test and maintenance. These ground rules have been extracted from EPRI TR-100381 (EPRI 1992) and from the experience of PRA vendors and NRC data analysts.

1. A maintenance event must result in the component not being capable of performing its function, as modeled in the PRA, in order to contribute to the component or train unavailability. For example, maintenance performed on a normally open MOV (that is required to stay open during its mission time) with the valve locked in the open position is not an event of interest. Similarly, a maintenance event involving some electrical repairs on an MOV that do not necessitate moving it from the position required for successful system operation is also not an event of interest. However, in either case, if the valve were required to close for any reason, then both events would be of interest.

2. Some testing procedures may result in component, train, or system unavailability. For example, a full flow test of a system through a test path could require that a normally closed injection valve be disabled in order to prevent inadvertent injection. The injection valve would be unavailable during the test period. However, systems often have logic which would actuate the system even if it was being tested. In this situation, there would be no system unavailability due to the test. A review of testing procedures coupled with knowledge of system actuation logic is required to determine if testing can result in component, train, or system unavailability.

3. If a maintenance report indicates that one or more trains of front line systems are unavailable due to maintenance activities of a support system, the unavailability is associated only with the support system.

4. If while performing maintenance on a support system, maintenance is also performed on the front line system it supports, the unavailability of the front line system should be counted if the two maintenance activities are not always performed together.

5. If an unavailability on one component is actually due to maintenance activity on another component that is included in the PRA model, the unavailability is associated with the second component only. For example, a declared unavailability of a pump due to maintenance on a room cooler should be included only as a maintenance on the room cooler if the dependence of the pump on the room cooler was modeled explicitly. As another example, if the maintenance results in the unavailability of a source of suction to a pump (e.g., maintenance on a supply tank), then it is better to model this as an unavailability of the source rather than the pump. Assigning the event to the source unavailability is absolutely required if the source is shared with other pumps. In general, maintenance unavailability should be allocated consistent with the component boundaries and system modeling.

6. There may be events where the unavailability of a component in a system model is due to maintenance on a component that is not included in any system model. In such cases, the event should be included as an unavailability of all the modeled components removed from service. For example, the contribution of maintenance on a drain valve for a pump train will likely not be modeled in the PRA but should be included as a contributor to the unavailability of the entire pump train since it would likely result in isolation of the train.

7. Coincident outage times for redundant equipment (both intra- and inter-system) should reflect actual plant experience. For some systems, the available redundancy may be higher than that limited by technical specifications. In this case, maintenance may be performed on two out of three trains at the same time. The modeling of dual component maintenance events in the PRA should be consistent with the actual plant experience. Note that because of the allowed outage time limitations in technical specifications, the maintenance unavailability may be lower when two trains are taken out for maintenance.

8. The maintenance data at the plant most likely will contain planned and forced maintenance. Most of the maintenance events will be forced type. If the PRA models the two types of maintenance separately and it is possible to distinguish between the two types in the data, these should be recorded separately.

9. In some cases, more than one maintenance activity may be recorded on an event report. When this occurs, each separate maintenance activity must be considered at the highest possible component level. For example, if the suction or discharge valve of a pump requires maintenance, the pump would be tagged out for the duration of the work. As previously discussed, the maintenance unavailability should be associated with the valve. If during this maintenance outage, some minor maintenance was performed on the pump, than the entire maintenance outage can be recorded as a pump maintenance event. The duration of the maintenance would be the time between when the first component is tagged out and when the last component is tagged in.

   However, if the maintenance unavailability is being modeled in the PRA at the train level, all maintenance activities on any component are included. In this situation, each maintenance event on any component in the train is included. If multiple components are tagged out during the maintenance event, the duration of the maintenance would be the time between when the first component is tagged out and when the last component is tagged in.

10. Functional dependencies represented in the PRA models must be considered in the allocation of maintenance events. For example, if a chilled water pump is taken out for maintenance, together with an HVAC chiller that it supports, only the chilled water pump is counted as being unavailable for maintenance. The functional dependency between the two components in the PRA model will account for the chiller being unavailable when the chilled water pump is under maintenance.

11. The cold shutdown periods in the time window over which data are being collected should be defined. The maintenance performed during shutdown is not included in the determination of component unavailability during power operation.

12. Special attention is required when allocating maintenance events for systems or components shared between units at a site. The technical specifications pertaining to shared systems can be different depending on the status of both units. The PRA model may include basic events to account for the dependence of the system unavailability on the mode of operation for each unit. In such cases, the maintenance events should be allocated to match those event definitions.

5.2.4 Component Exposure Evaluation

The data identification and allocation process discussed in the previous sections results in the identification of the number of events associated with each component failure mode. To generate component failure probabilities and rates, it is also necessary to estimate the operational exposure of the components. The term “exposure” refers to the amount of component operating time when considering failure rates and to the number of demands (or cycles) when considering failure probabilities.

Exposure data are normally developed by reviewing plant documents; e.g., test procedures and the knowledge of component function (standby, normally operating, etc.), and systems lineup. In some cases, an operation time meter provides information about the cumulative hours of operation of a component.

Development of exposure data involves many judgments and assumptions. The guidance provided in this section sometimes leads to an approximate value for the exposure data, which may differ substantially from the actual experience. Although typically the range of uncertainties associated with the exposure data are much smaller than those for the failure data, there may be cases where the combined effect of uncertainty about the exposure and failure has a significant impact on the estimate of the failure rate or probability. The issue of uncertainty in the data (both in the failure and exposure data) is addressed in Section 6.1.2.2 of this handbook.

The following sections outline the process for estimating the number of demands and the operating time for each component. Much of this guidance is taken from EPRI TR-100381 (EPRI 1992).

5.2.4.1 Time-Related Exposures

The operating or exposure time for a component is dependent upon whether the component is normally operating or is in standby. For components that are required to continuously operate during a particular plant mode, the operating time can be easily established by directly relating it to the time spent in that plant mode.

Some plant systems, sometimes called alternating or intermittently operated systems, have multiple redundant trains where only a subset of those trains are required to operate at any one time. A standard practice at nuclear power plants is to alternate the trains that are

operating and in standby at specified intervals. The times of operation and changeover from one train to another are typically recorded in the control room or some other log book. However, since the pumps in different trains of a system are usually grouped together for data analysis, it is not necessary to have an accurate log of how long an individual pump was in operation. Instead, it is only necessary to evaluate the exposure time for the pumps as a group. For example, if two of three pumps are normally operating in a particular plant mode, the total operating time for that pump group is twice the calendar time spent by the plant in that mode.

For a component in a standby system, the operating time is generally given by the time the system is operated during testing. Note that an important criterion for including test data when evaluating both the failure and exposure data is that the test should mimic the component operation that would be required in an unplanned demand. The testing period may be recorded in control room logs or other logs. The operating time during testing for a population of components may also be estimated by summing the product of the component population, test frequency, and test duration for each test during the period where failure data was collected. It should be noted that for most plants, and most components, the cumulative run time during testing is relatively short.

Some systems that are in standby during normal power operation are also used during other modes of operation. For example, the residual heat removal (RHR) system in both BWRs and PWRs is used during shutdown. Similarly, a standby system may be used during power operation for a special purpose. For example, the RHR system in a BWR may be used to increase or decrease the suppression pool level. Thus the operating times during these modes of operation should be included, in addition to the run times during testing, if any failures during these modes are pertinent to the safety function of the system (e.g., the entire RHR pump operating history may be pertinent since the pump must operate when the RHR system is used to respond to an accident). In such situations, the times of startup and shutdown of the standby system may be recorded in the control room logs. Alternatively, if the component is required to continuously operate during shutdown , the operating time can be easily established by directly relating it to the time spent in that plant mode.

5.2.4.2 Demand-Related Exposures

To evaluate the probability of the failure of a component to start or change states, the number of demands experienced by the component must be evaluated. Although this would seem to be a simple process, in practice the number of demands is often one of the most difficult parameters to calculate accurately. Component demands from all contributors should be included. This can include contributions from testing, automatic and manual actuations, and corrective maintenance. The methods of calculating the number of demands from each of these types of demands are explained below.

5.2.4.2.1 Test Demands

Periodic testing is an important source of demands for components in standby systems. The surveillance testing and required frequency for the plant is performed in accordance with the technical specifications. However, some plants may choose to perform testing more frequently than required by the technical specifications.

An important criterion for including test data in evaluating both the failure and exposure data is that the test should mimic the component operation that would be required in an unplanned demand.

Surveillance procedures identify the components that must change state at each test. For each surveillance test pertinent to the system, it is important to identify which components are operated, the unavailability of the system during the test (if applicable), and the frequency and duration of the test. A functional test of a pump often requires the operation of valves as well as the pump and is an important source of information on valve demands. Neglecting demands on components from tests on other components can lead to a significant underestimation of the total number of demands. The number of test demands for individual components may be determined from the actual number of tests as recorded in a control room or test logs or be estimated based on the test frequencies.

It should be noted that the test may not be a valid test for all the components within the component boundary. For example: the automatic initiation portion of a component circuit will not be tested during a test where the component is manually initiated. For components such as diesel generators, tests which start the engine, but do not close the breaker onto the bus are not true tests of the capability of the diesel generator to provide the necessary load. Note that if there is a subcomponent that is included in a component's boundary which is not tested along with the rest of the component, it is desirable to analyze it as a separate component.

5.2.4.2.2 Automatic and Manual Initiation

Actual unplanned demands on components should be included in the demand count. For standby safety system components, some unplanned demands can be traced back to the occurrence of automatic initiation signals (both actual and spurious signals). These signals include emergency core cooling system (ECCS) initiating signals, turbine trip signals, losses of offsite power, and reactor scrams. Different groups of component may be initiated by different signals or sets of signals, depending on the functions and the system they are in. Information on the components that can be initiated by each signal can be identified through knowledge of the plant. For example, all low-pressure ECCS pumps in a BWR could be initiated by an ECCS signal but the motor-operated valves in the ECCS injection paths would require an additional low vessel pressure signal before they would open. Information on the historical number of occurrences of actual or spurious signals should be available from the plant records such as the monthly operating reports or control room logs.

In addition, manual actuation of systems or components may occur during plant operation. Two examples cited above in the discussion of operating time contributors are also pertinent here. The first is the case where alternating trains are placed in operation and standby. The act of switching operating trains results in demands on components. The second case involves the use of standby systems to perform special functions. For example, the RHR system in a BWR may be used to increase or decrease the suppression pool level. These special uses also result in component demands. In both cases, the times of startup and shutdown of the standby system may be recorded in the control room or other types of logs.

Finally, manual actuation of systems to respond to adverse plant conditions is another source of unplanned demands that needs to be accounted for in the exposure evaluation. The occurrences of such demands are generally recorded in LERs, control room logs, and monthly operating reports.

5.2.4.2.3 Corrective Maintenance

Maintenance can result in demands on components in several ways. Before the maintenance activities are begun, the operating and maintenance staff make the maintenance action safe for both personnel and the system by disabling and tagging out appropriate components. This then requires some components to change state resulting in a demand.

In many instances, demands are placed on components that are not the subject of the corrective maintenance. The most obvious demands occur when a component is returned to service. Before restoring the component to service following maintenance, a complete functional checkout is usually performed on the component and other components in the functional loop. The number of demands on the components resulting from corrective maintenance is obtained from the number of maintenance acts on specific components and an identification of what other components may have to change state to complete the functional test. **Note that per the guidance in the ASME PRA Standard (ASME 2002), demands from post-maintenance testing should be excluded from the exposure evaluation for the component under maintenance.**

Another example of a demand resulting from maintenance involves testing of redundant trains. If equipment fails in some systems, the technical specifications may require that redundant components be checked for operability before maintenance to ensure that they are available for service. In many cases, an increased frequency of surveillance testing of such redundant components is required. A typical example of this is reflected in the technical specifications for emergency diesel generators. These demands need to be included in the data analysis.

As indicated in the discussions presented above, development of exposure data involves many judgments and assumptions. Although typically the magnitude of error or the range of uncertainties associated with the exposure data are small compared with those of the failure data, there are cases where the combined effect of uncertainty about the exposure and failure has a significant impact on the estimate of the failure rate. The data analyst should consider some level of uncertainty in using such estimates.

5.2.5 Determination of Unavailable Time

Following the identification of the maintenance events contributing to the unavailability of a component, train, or system, the time the component is unavailable during each event is determined. The unavailability time is the time between when the component is removed from service until it is actually restored to service. In many cases, maintenance work orders will provide this information by identifying one or more tag-ins and tag-outs for equipment with the date and time of day that both occur. Using these times to determine the unavailability time may be a little conservative because the repair may be completed before the component is declared tagged in.

Some maintenance work orders may contain multiple tag-outs and tag-ins for a given component. If the component was operable between these periods, than the unavailability is the sum of the individual unavailability times for each period. However, if the component was inoperable between the periods, than the unavailability time starts at the first tag-out and ends at the last tag-out.

Unfortunately, the actual time of unavailability may not be recorded in maintenance work order forms. In many cases, the time recorded may reflect a prior estimate of how long the maintenance activity will take, may represent the man-hours taken to complete the task rather than calendar time, or may include time to complete paperwork.

When the unavailability time is not specified in a maintenance work order, other plant documents should be examined for that information. Maintenance activity information may be recorded in other documents such as operator logs or component operating logs. For example, a maintenance activity on a safety-related component will start the clock for a limiting condition of operation (LCO) specified in the technical specifications, and this should be recorded in some place, usually the control room log. The time when the function is restored should also be recorded. Unfortunately, not all maintenance events result in an LCO and thus timing information may not be available.

When reliable estimates of the start and finish times for a maintenance event are not available, one recourse is to ask plant maintenance and operations staff to provide estimates of the ranges in the unavailable time per maintenance act for the components. Another recourse is to use data provided from some maintenance events to estimate the unavailability for other events.

5.3 Recovery Event Data

In PRA, there is a clear distinction between actions to repair components or systems and actions to recover components or systems. Recovery actions involve the use of alternate equipment or means to perform a function when primary equipment fails, or the use of alternate means to utilize equipment that has not responded as required. Examples of recovery actions include opening doors to promote room cooling when an HVAC system fails, recovering grid-related losses of offsite power by rerouting power, manually initiating a system when the automatic actuation signal fails, bypassing trip logic using jumper cables, and using a handwheel to manually open an MOV when the motor fails to operate. Repair actions involve the actual repair of the mechanism which caused a component or system to fail. Examples of repair actions include repairing weather-related losses of offsite power, repair of a pump that failed to start, or replacement of a failed circuit breaker.

PRA models typically include a number of recovery actions of the type identified above. However, because recovery actions can involve complicated actions that are governed by procedures, most are typically evaluated using HRA methods. A general exception is the treatment of offsite power recovery where the required recovery actions are often not within the jurisdiction of the plant personnel. Thus, offsite power recovery data is collected and reduced for use in PRAs.

The repair of components is generally not modeled in PRAs since:

- the time available to repair most components is generally too limited (i.e., core damage would occur before the repair is completed),
- because repair is an action that is not always governed by procedures and thus difficult to justify,
- the availability of spare parts can not always be certain, and
- because abnormal procedures generally direct operators to use alternative equipment as a first priority.

There are always exceptions to these general observations. For example, the replacement of fuses is an action identified in some fire abnormal procedures and can be accomplished rather quickly since spare fuses are available. As with a recovery action, either an HRA or data reduction approach could be utilized to generate a failure probability for a repair action.

The modeling of recovery and repair actions in PRA reflects the need to accomplish the action within some time frame (e.g., before core damage occurs). Thus, the collected data must include both the time of failure and recovery to be utilized in the PRA. This section provides guidance on the process for collecting and reducing recovery and repair data. A description of the type of data that is reviewed in this effort and guidelines for allocating that data.

5.3.1 Recovery Data Identification

Recovery and repair information can generally be extracted from maintenance records and LERs that

identify component and system failures. Thus, the evaluation of recovery and repair information is an offshoot of the component failure data review. In general, only data from actual component and system demands should be included in the recovery/repair data evaluation. When failures occur during actual demands, operators should be strongly motivated to try to recover the component or system.

However, if a component or system fails to start during a surveillance test, the need for repair is not as pressing and thus not reflective of accident conditions. For this reason, recovery and repair information for failures during surveillance tests should be excluded from recovery/repair probability evaluation.

5.3.2 Recovery Data Allocation

Since component recovery data evaluation should be performed in conjunction with the component data allocation, the general rules provided in Section 5.2.3 apply. In addition, the following guidelines are provided to address allocating recovery data for other events modeled in the PRA (e.g., restoring offsite power or reopening main steam isolation valves):

1. Only failures during actual demands are included. Failures during surveillance tests are excluded as being nonrepresentative of accident conditions. For the failures during actual demands, the data analyst should assess whether the recovery/repair action was performed under similar stresses that would occur under accident conditions. Atypical events should be eliminated or considered to be sources of uncertainty.

2. For each failure event, the recovery/repair time is the time between when the failure first occurs and the time when it is returned to service. Using these times ensures that the time of the failure, the time required to recognize it has occurred, the time to obtain spare parts if required, the actual time to repair the component or system, and the time to return the component to service are reflected in the recovery/repair time. Events that do not include either time should be excluded from the evaluation.

3. Recovery information on systems or components resulting from an initiating event can be extracted from LERs or scram reports. For example, reopening MSIVs after their consequential closure (i.e., they are signaled to close following some other failure) may be included in a PRA for some initiators. The recovery time for such events are evaluated from the time the initial failure occurs leading to MSIV closure to until the closure signal is removed (by either fixing the original failure or by bypassing the signal) and the MSIVs in one hot leg are reopened. The time to perform other actions that may be required to maintain the MSIVs open (e.g., starting vacuum pumps) are also included in establishing the recovery time.

4. Recovery information on systems or components causing an initiating event can also be extracted from LERs or scram reports. For example, the time to recover offsite power initiating events can be extracted from LERs. However, LERs should also be searched for occurrences of offsite power failure following other initiating events. Recovery information should also be extracted for these events.

6. PARAMETER ESTIMATION AND MODEL VALIDATION

6.1 Overview

6.1.1 Chapter Contents

This chapter is the heart of the parameter-estimation portion of this handbook. Section 6.1 gives an important discussion of Bayesian and frequentist inference, and also a brief discussion of some topics outside the scope of the handbook. The rest of Chapter 6 presents statistical techniques for analyzing data for various parameters. Sections 6.2 through 6.7 cover exactly the same types of data as Sections 2.2 through 2.6, in the same order. The two kinds of failure to start in Section 2.3 are split into two sections here, 6.3 and 6.4. The three most extensive and fundamental sections are 6.2 (initiating events), 6.3 (failures on demand), and 6.6 (recovery times and other durations). The remaining sections draw on material from these three. Figure 6.1 shows the contents in a schematic way, with arrows indicating the logical dependencies. For example, Section 6.4 uses material presented in Sections 6.2 and 6.3.

Each section considers both parameter estimation and model validation. These two topics are considered together because checking the assumptions of the model (model validation) is a necessary part of any analysis. Separating the model validation from the parameter estimation might give the erroneous impression that it is all right to estimate parameters without checking the assumptions, or that the checks can be performed as an afterthought.

Under parameter estimation, both Bayesian and frequentist methods are presented. Under model validation, both graphical methods and formal statistical tests are given.

Much thought was given to the order of presentation: do we present the Bayesian estimates first or the frequentist estimates? In Chapter 6, the frequentist estimates are typically given first, not because they are more important or more highly recommended, but only because the frequentist point estimates are very simple, the simplest most natural estimates that someone might try. We cover them quickly before moving on to the more sophisticated Bayesian estimates. In the cases where the frequentist estimates are not simple (such as certain distribution models for durations), Bayesian estimation is discussed first.

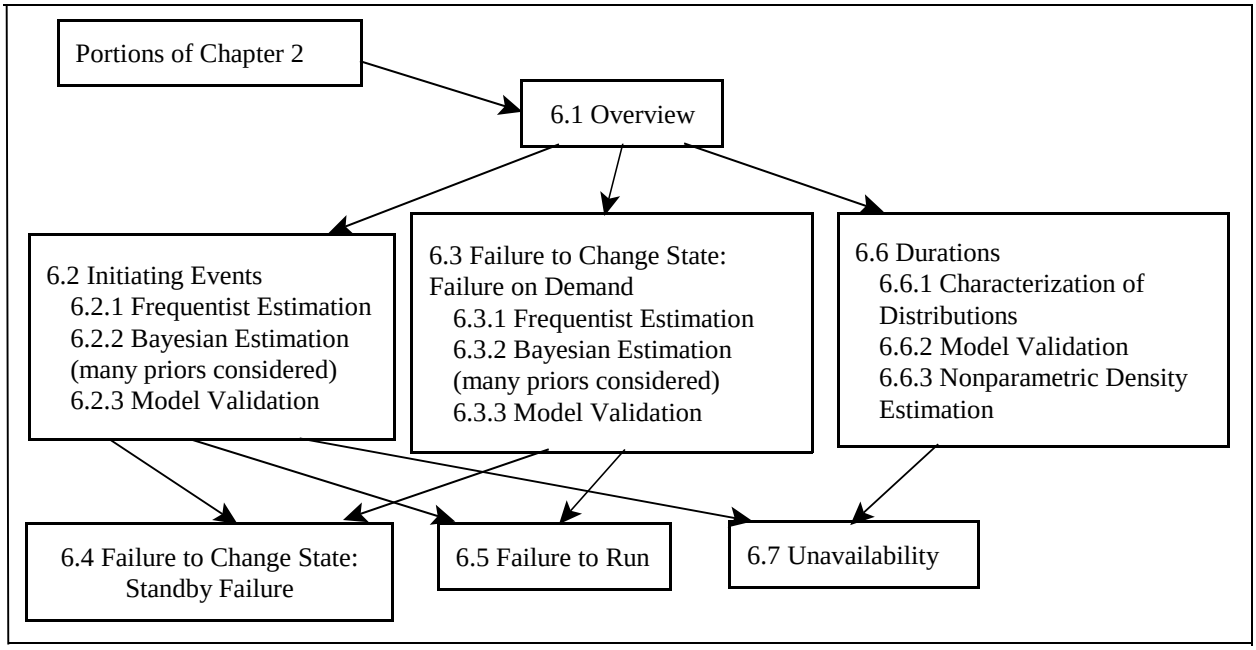


Figure 6.1. Schematic outline of Chapter 6.

As in much of this handbook, general explanations are given in Roman typeface, with **boldface** used for new terms where they are introduced or defined. Arial font is used for examples, and for any extended discussion that applies only to a particular example.

6.1.2 Bayesian and Frequentist Inference

Probabilistic risk assessment (PRA) analyzes accident sequences in terms of initiating events, basic events, and occasionally recovery events.

This handbook is concerned with estimating the frequencies of initiating events, the probabilities of basic events, and the distributions of recovery times and other durations. These estimates are propagated through logical relations to produce an estimated frequency of the undesirable end state, such as core damage. Moreover, the uncertainties in the parameter estimates must be quantified, and this must be done in a way that allows the uncertainty in the final estimate to be quantified.

Two approaches to estimating parameters are the Bayesian method and the frequentist, or classical, method. The two approaches are summarized here, and also in Appendix B.

Both approaches use probability distributions to describe the behavior of random outcomes, such as a random number of initiating events or a random number of failures to start. The two approaches differ in the way they treat uncertainty of unknown parameters.

In the Bayesian setting, probability is a measure of uncertainty, a quantification of degree of belief. The Bayesian methodology is used to modify uncertainty in a logically coherent way, so that "degree of belief" is rational, not merely personal opinion. In this methodology, each unknown parameter is assigned an initial **prior** probability distribution. This does not mean that the parameter varies randomly, but only that it is unknown, with the probability distribution modeling belief concerning the true value. Based on data, the analyst's prior belief about the parameter is updated, using Bayes' Theorem. The final inference statement uses the **posterior** distribution of the parameter to quantify the final uncertainty about the parameter. It is conditional on the observed data. Siu and Kelly (1998) give a simple but thorough introduction to Bayesian estimation in the PRA context.

The frequentist approach is quite different. The probability of a random event is defined as the long-term fraction of times that the event would occur, in a large number of trials. Probabilities are used only for random quantities, the possible data values. Probability distributions are never used to describe parameters, because the parameters are not random. When quantifying uncertainty in an estimate, a frequentist asks questions such as, "Under similar conditions, what other data sets might have been generated? From data set to data set, how much variation would be seen in the parameter estimate? For any one data set, how far might the estimated parameter be from the true parameter?" Any prior or external information about the parameter value is ignored.

Statisticians have argued vigorously over which approach is preferable. When estimating parameters for PRA, the Bayesian approach clearly works better, for two reasons. First, data from reliable equipment are typically sparse, with few or even zero observed failures. In such cases, it is reasonable to draw on other sources of information. The Bayesian approach provides a mechanism for incorporating such information as prior belief. Second, the Bayesian framework allows straightforward propagation of basic event uncertainties through a logical model, to produce an uncertainty on the frequency of the undesirable end state. To do this, it assigns a probability distribution to each of the unknown parameters, draws a random sample from each, and constructs the corresponding sample for the frequency of the undesirable end state. The frequentist approach cannot handle such complicated propagation of uncertainties except by rough approximations.

Frequentist methods have their uses, however, even in PRA. Box (1980) writes "sampling theory [the frequentist approach] is needed for exploration and ultimate *criticism* of an entertained model in the light of current data, while Bayes' theory is needed for *estimation* of parameters conditional on the adequacy of the entertained model." This viewpoint agrees with current PRA practice. The primary use of the frequentist approach is in preliminary examination of the data, to check the correctness of model assumptions, and to decide which model to use. For example, frequentist methods can help the analyst decide whether data sets may be pooled or whether a trend is present. Goodness-of-fit tests and calculation of statistical significance are commonly used frequentist tools in this context. Then Bayesian methods are used for estimating the parameters. In addition, frequentist estimates are often simpler to calculate than Bayesian estimates, and therefore are useful for rough approximate calculations.

Table 6.1 summarizes the above points.

Table 6.1 Comparison of Bayesian and frequentist approaches in PRA.

	Frequentist	Bayesian
Interpretation of probability	Long-term frequency after many hypothetical repetitions.	Measure of uncertainty, quantification of degree of belief.
Unknown parameter	Constant, fixed.	Constant, but assigned probability distribution, measuring current state of belief.
Data	Random (before being observed).	Random for intermediate calculations. Fixed (after being observed) for the final conclusions.
Typical estimators	Maximum likelihood estimator (MLE), confidence interval.	Bayes posterior mean, credible interval.
Interpretation of 90% interval for a parameter	If many data sets are generated, 90% of the resulting confidence intervals will contain the true parameter. We do not know if our interval is one of the unlucky ones.	We believe, and would give 9 to 1 odds in a wager, that the parameter is in the interval.
Primary uses in PRA	1. Check model assumptions. 2. Provide quick estimates, without work of determining and justifying prior distribution.	1. Incorporate evidence from various sources, as prior distribution. 2. Propagate uncertainties through fault-tree and event-tree models.

6.1.3 Uncertainties Other Than Parametric Uncertainty

The above discussion might suggest that uncertainty in the value of parameters is the only uncertainty there is. That is not the case. Parameter uncertainty, stemming from having only a relatively small set of randomly generated data, is the simplest uncertainty to address. It is the primary uncertainty considered in this handbook of parameter estimation. However, the following kinds of uncertainty can also be considered. Because these subsections discuss material that is outside the scope of the handbook, first-time readers may wish to skip immediately to Section 6.2.

6.1.3.1 Uncertainty from Nonrepresentativeness of the Data Sources

One issue to consider is that the data come from settings that do not perfectly match the problem of interest. In general, this is a difficult issue. For example, suppose one situation is of interest, but the data come from equipment with a different manufacturer or different design, or from equipment operated under different conditions, or maintained with different practices. Then it is difficult to quantify the relationship between the data and the problem of interest. Engineering judgment is used, and to be conservative the uncertainty distribution is often assigned a larger variance than the data alone would call for.

One tractable case is uncertainty of the value of a parameter for one data source (such as one nuclear power plant), when data are available from many similar but not identical data sources (other nuclear power plants). This case can be formulated in terms of a hierarchical model, and analyzed by empirical Bayes or hierarchical Bayes methods, as discussed in Chapter 8 of this handbook.

6.1.3.2 Uncertainty in the Data Counts Themselves

There can be uncertainty in the data counts themselves. For example: it may be unclear whether a particular event should be counted as a failure, or the number of demands may not be known exactly. A Bayesian

method for dealing with uncertainty in PRA data was first proposed by Siu and Apostolakis (1984, 1986), and has been used by several authors, including Mosleh (1986), Mosleh et al. (1988, Section 3.3.4.4), and Martz and Picard (1995). As outlined by Atwood and Gentillon (1996), uncertainty in classifying the data yields a number of possible data sets, each of which can be assigned a subjective probability. The simple approach is to use an "average" data set, a "best estimate" of the data, and analyze it. The uncertainty in the data is ignored, lost, at that point. A better approach is to analyze each data set, and combine the results. Each analysis produces a Bayesian distribution for the unknown parameter(s), and the final result is a mixture of these distributions. This approach includes the data uncertainty in the analysis, and results in wider uncertainty intervals than the simple approach. The two approaches are diagramed in Figure 6.2.

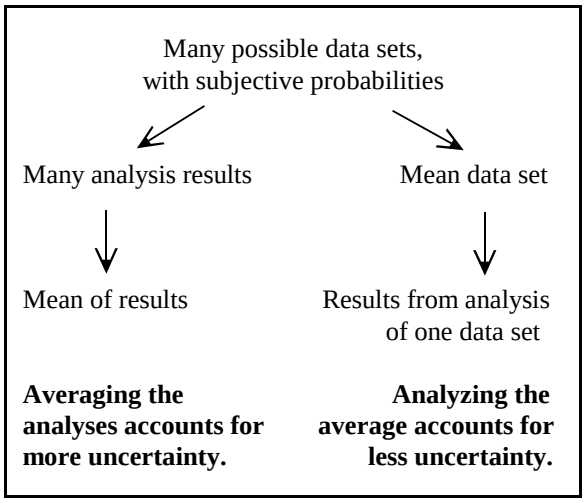


Figure 6.2 Two possible analysis paths for uncertain data.

Data uncertainty has become the subject of recent journal articles, such as the by Martz and Hamada (2003), who develop a fully Bayesian method. Also, this topic is closely related to a statistical technique called "multiple imputation" (see Rubin 1996), in which a moderate number of data sets are randomly generated and then treated according to the left path in Figure 6.2. Further treatment of this topic is beyond the scope of this handbook, but the reader can find additional guidance in the references cited above.

6.1.3.3 Uncertainty in the Correct Model to Use

There can be uncertainty in which probability model to use. For example, there may be a slight trend, but it is borderline. Should a trend be modeled? Chapters 6 and 7 of this handbook discuss model validation extensively. However, model validation, which concludes that the model is either "adequate" or "not adequate," is only a first step toward addressing this issue.

A more ambitious approach would be to quantify the degree of belief in each of a number of models, and propagate uncertainty in the models into the overall conclusions. This approach can use the predictions of various models as evidence in a formal Bayesian estimation procedure. See Mosleh et al. (1994) for a number of thoughtful papers on the definition and treatment of model uncertainties in the context of PRA applications. The topic is also discussed and debated in a tutorial article by Hoeting et al. (1999). Bernardo and Smith (1994) also work out this approach in their Chapter 6 on "remodelling." Drougett (1999) includes a discussion on the role of information concerning the models themselves (for example, their structure and past performance) in the estimation process.

Further consideration of such issues is beyond the scope of this handbook. The parameter uncertainties given here all assume that the model is a perfect description of the real world.

6.2 Initiating Events

This section and Section 6.3 are fundamental. The methods introduced here are used throughout the rest of the handbook. The most important topics for a first-time reader are:

- Maximum likelihood estimation (6.2.1.1),
- Bayesian estimation, especially with a discrete prior or a conjugate prior (6.2.1-6.2.2.5), and
- Model validation, especially using graphical tools (portions of 6.2.3).

Initiating events here use the broad definition of the examples in Section 2.2, events that occur randomly in time and that initiate a quick response to restore the system to normal.

The event frequency is denoted δ, with units events per unit time. The data consist of x observed events in time t, where x is an integer ≥ 0 and t is some time > 0. Note, t is considered nonrandom, and x is randomly generated. This can be expressed using the notation given in Appendix A, with upper case letters denoting random variables and lower case letters denoting numbers. Before data had been generated, the random number of initiating events would have been denoted by X. For any particular number x, the probability of x initiating events in time t is

$$\Pr(X = x) = e^{-\delta t}(\delta t)^x / x! \; . \qquad (6.1)$$

This formula for the Poisson distribution is a restatement of Equation 2.1, and will be used throughout this section.

The methods of parameter estimation will be illustrated by the following hypothetical data set.

Example 6.1 Initiating events with loss of heat sink.

In the last six years (during which the reactor was critical for 42800 hr.) a hypothetical PWR has had one initiating event that involved a loss of heat sink. The parameter to estimate is λ, the frequency of such events while the reactor is critical.

6.2.1 Frequentist or Classical Estimation

As explained in Section 6.1, Bayesian estimation methods are more important in PRA, but the classical estimator has a simpler form. Also, the comparison among estimators flows somewhat better if the short presentation of frequentist estimators precedes the lengthier presentation of Bayesian estimators. For these reasons, frequentist methods are given first in this section.

6.2.1.1 Point Estimate

The most commonly used frequentist estimate is the **maximum likelihood estimate** (MLE). It is found by taking the **likelihood**, given by Equation 6.1, and treating it as a function of λ. The value of λ that maximizes the likelihood is called the MLE. It can be shown (as a calculus exercise) that the maximum likelihood estimate (MLE) of λ is

$$\hat{\lambda} = x / t \ . \qquad (6.2)$$

This formula is simple and intuitively natural: the observed number of events divided by the observed time period. This simplicity is part of the appeal of the MLE. The hat notation is used to indicate that the MLE is an estimate calculated from the data, not the true, unknown λ.

Example 6.1 has x = 1 and t = 42800 hrs. The likelihood is plotted on Figure 6.3 as a function of λ.

The likelihood function is maximized when λ = 1/42800 = 2.3E-5. Therefore, the estimated event rate for the plant is

$\hat{\lambda}$ = 1/42800 = 2.3E!5 events per critical-hour .

Converting the hours to 42800/8760 = 4.89 critical-years yields

$\hat{\lambda}$ = 1/4.89 = 0.20 events per critical-year.

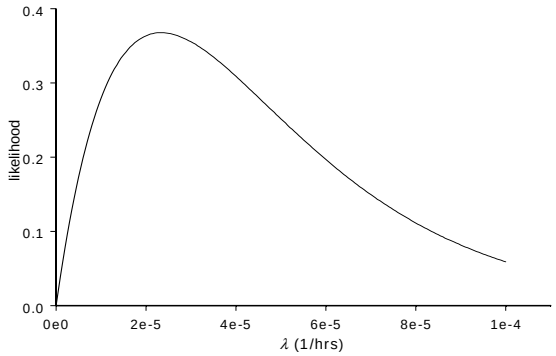


Figure 6.3 Likelihood as a function of λ, for data of Example 6.1.

In the above example, and in general throughout this handbook, the final answer is presented with few significant digits. This reflects the uncertainty inherent in all estimates. Indeed, sometimes not even the first significant digit is known precisely. During intermediate calculations, however, more significant digits will be shown, and used. This prevents roundoff errors from accumulating during the calculations.

It is also possible to combine, or **pool**, data from several independent processes, each having the same rate λ. In particular, suppose that the ith Poisson process is observed for time t_i, yielding the observed count x_i. The total number of event occurrences is $x = \Sigma_i x_i$, where the sum is taken over all of the processes, and the exposure time is $t = \Sigma_i t_i$. The rate λ is estimated by

$$\hat{\lambda} = x / t = \Sigma_i x_i \ / \ \Sigma_i t_i \ .$$

For example, if counts obtained for different years are used to estimate the rate, the estimate is the ratio of the total count to the total exposure time during these years.

6.2.1.2 Standard Deviation of the Estimator

The event count is random. In other words, if an identical plant could be observed during the same years, a different number of events might occur due to randomness. Similarly, the same plant might yield a different count over a different six-year period.

Because the event count is random, the estimator is also random, and the estimate is simply the observed value for this plant during these years. Note the distinction in the terms: an **estimator** is a random variable, and an **estimate** is the particular value of the estimator after the data have been generated.

For a Poisson distributed random variable X, the mean and variance are the same, $E(X) = \text{var}(X) = \lambda t$, as stated in Appendix A.6.2. Consequently, the standard deviation of X is $(\lambda t)^{1/2}$, and the estimated standard deviation of the estimator $\hat{\lambda} = X/t$ is

$$(\hat{\lambda} t)^{1/2} / t = (\hat{\lambda} / t)^{1/2} = x^{1/2} / t \quad .$$

The estimated standard deviation of $\hat{\lambda}$ is also called the **standard error** for λ.

Thus, the standard error for λ in Example 6.1 is 1/4.89 = 0.20 events per reactor-year.

A standard error is sometimes used for quick approximations when the data set is large. In that case, the MLE is approximately normal, and an approximate 95% confidence interval is given by MLE ± 2×(standard error). This approximation holds for maximum likelihood estimation of virtually any parameter, when the date set is large. For event frequencies, however, the following exact confidence interval can be found.

6.2.1.3 Confidence Interval for λ

Frequentist estimation is presented before Bayesian estimation because the MLE is so simple, simpler in form than the Bayes estimates. The same cannot be said for confidence intervals; the confidence-interval formulas are somewhat more complicated than the formulas for Bayesian interval estimates, and the interpretation of confidence intervals is more subtle. Confidence intervals are used in two ways in this handbook: they give a standard of comparison, when Bayes credible intervals are found based on so-called noninformative priors, and they can be used (but are not required) in some plots for validating model assumptions. Therefore, readers may wish to skim the present section quickly on the first reading.

The confidence interval is given in many reference books, such as Johnson, Kotz, and Kemp (1992, Sec. 7.3), Bain and Engelhardt (1992, Section 11.4), or Martz and Waller (1991, Table 4.4). It is based on the chi-squared (or in symbols, χ^2) distribution, which is tabulated in Appendix C, and which can be found easily by many software packages. As used below, $\chi^2_p(d)$ is the pth quantile, or $(100p)$th percentile, of the chi-squared distribution with d degrees of freedom. Do not misread $\chi^2_p(d)$ as involving multiplication.

For a $(1 - \alpha)$ confidence interval, or equivalently a $100(1 - \alpha)\%$ confidence interval, the lower limit is

$$\lambda_{\text{conf},\,\alpha/2} = \frac{\chi^2_{\alpha/2}(2x)}{2t}$$

If $x = 0$, this formula is undefined, but then simply set $\lambda_{\text{conf},\,\alpha/2} = 0$.

Similarly, the upper limit is

$$\lambda_{\text{conf},\,1-\alpha/2} = \frac{\chi^2_{1-\alpha/2}(2x+2)}{2t} .$$

Notice that an upper confidence limit is defined in the case $x = 0$. It is reasonable that observing no occurrences of the event would provide some information about how large λ might be, but not about how small it might be.

The above formulas are in terms of α. Setting $\alpha = 0.1$, for example, gives the formulas for a 90% confidence interval. These formulas involves the 5th percentile of a chi-squared distribution with $2x$ degrees of freedom, and the 95th percentile of a chi-squared distribution with $(2x+2)$ degrees of freedom.

The resulting confidence interval is conservative in the sense that the actual confidence level is no smaller than the nominal level of $100(1 - \alpha)\%$, but it could be larger. This conservatism is inherent in confidence intervals based on discrete data.

In Example 6.1, Table C.2 shows that 90% confidence limits are

$$\lambda_{\text{conf},\,0.05} = \frac{\chi^2_{0.05}(2)}{2 \times 4.89} = \frac{0.103}{9.78} = 0.010$$

$$\lambda_{\text{conf},\,0.95} = \frac{\chi^2_{0.95}(4)}{2 \times 4.89} = \frac{9.488}{9.78} = 0.97$$

with units events per critical-year.

The interpretation of confidence intervals is given in Appendix B. This interpretation deserves emphasis, so we elaborate on the topic here. In the frequentist approach, θ is fixed and the data are random. Therefore, the maximum likelihood estimator and the confidence limits are all random. For most data sets the MLE, $\hat{\lambda}$, will be close to the true value of θ, and the confidence interval will contain θ. Sometimes, however, the MLE will be rather far from θ, and sometimes (less than 10% of the time) the 90% confidence interval will not contain θ. The procedure is good in the sense that most of the time it gives good answers, but the analyst never knows if the current data set is one of the unlucky ones.

To illustrate this, consider the following example with many hypothetical data sets from the same process.

Example 6.2 Confidence intervals from computer-generated data.

A computer was used to generate Poisson data, assuming an event rate θ = 1.2 events per year and assuming that 6 years were observed. Thus, the event count followed a Poisson distribution with mean θt = 7.2. This was repeated, and 40 event counts were generated in all. These may be interpreted as counts from 40 identical plants, each observed for 6 years, or from 40 possible six-year periods at the same plant.

Figure 6.4 shows that the first randomly generated event count was 10, the next was 5, the next was again 10, and so on. Some of the event counts were less than the long-term mean of 7.2, and some were greater. The maximum likelihood estimates of θ are plotted as dots in Figure 6.4. The corresponding 90% confidence intervals for θ are also plotted.

In Figure 6.4, the vertical dashed line shows the true value of θ, 1.2. Two of the 40 intervals (5%) are to the right of the true θ. These resulted from observing event counts of 14 and 16. One of the 40 intervals (2.5%) is to the left of the true θ. This interval was computed from an observed event count of two.

Ideally, the error rates should both be 5%. They are not, for two reasons. First, 40 is not a very large number, so the random data do not exactly follow the long-run averages. Second, confidence intervals with discrete data are inherently conservative: a 90% confidence interval is defined so that the probability of containing the true θ is *at least* 90%, and the error probabilities at each end are each *at most* 5%.

The data analyst will normally have data from just one plant for the six-year period. The resulting confidence interval will contain the true value of θ, unless the data happen to deviate greatly from the mean. Unfortunately, the analyst does not know when this has happened, only that it does not happen often.

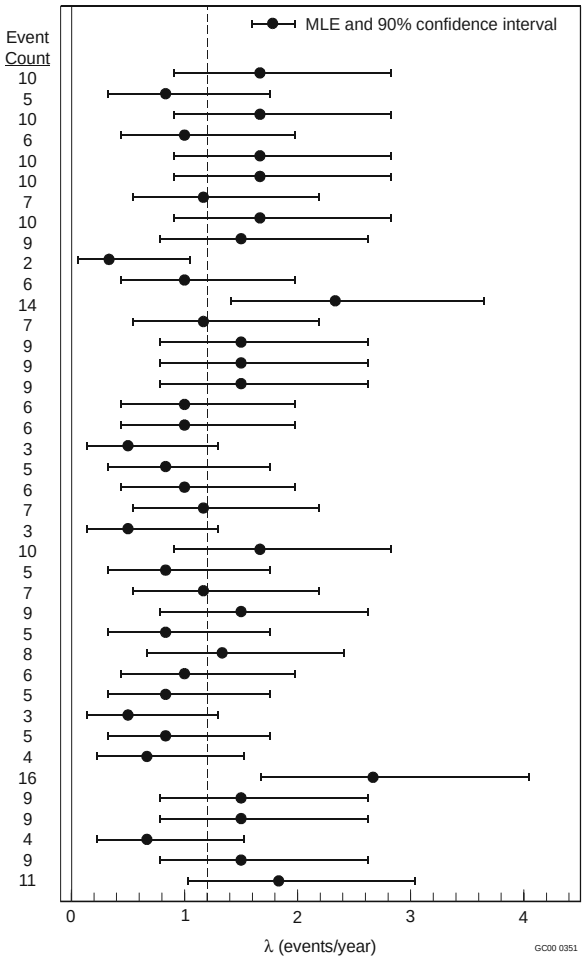


Figure 6.4 Confidence intervals from random data, all generated from the same process.

6.2.2 Bayesian Estimation

6.2.2.1 Overview

Bayesian estimation of θ involves several steps. The prior belief about θ is quantified by a probability distribution, the **prior distribution**. This distribution will be restricted to the positive real line, because θ must be positive, and it will assign the most probability to the values of θ that are deemed most plausible. The data are then collected, and the **likelihood function** is constructed. This is given by Equation 6.1 for initiating events. It is the probability of the observed data, written as a function of θ. Finally, the **posterior**

distribution is constructed, by combining the prior distribution and the likelihood function through Bayes' theorem. (For background on Bayes' theorem and Bayesian estimation, see Appendix A.5 and B.5.1.) This theorem says that

$f_{post}(\theta) \propto \text{likelihood}(\theta) \times f_{prior}(\theta)$.

Here, the symbol $\propto$ denotes "is proportional to." The posterior distribution shows the updated belief about the values of θ. It is a modification of the prior belief that accounts for the observed data.

Figure 6.5, adapted from a tutorial article by Siu and Kelly (1998), shows how the posterior distribution changes as the data set changes. The figure is based on a diffuse prior, and on three hypothetical data sets, with x = 1 event in t = 10,000 hours, x = 10 events in t = 100,000 hours, and x = 50 events in t = 500,000 hours, respectively. Note, each of these data sets has $\hat{\lambda} = x/t$ = 1.E!4 events per hour. The figure shows the prior distribution, and the three posterior distributions corresponding to the three data sets.

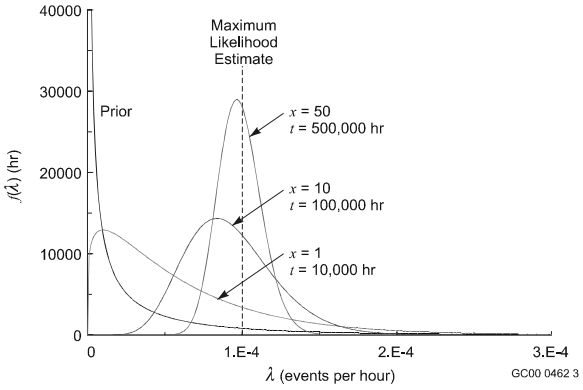


Figure 6.5 Prior distribution and posterior distributions corresponding to three hypothetical data sets.

For a small data set, the posterior distribution resembles the prior to some extent. As the data set becomes larger, several patterns are evident:

- the posterior distribution departs more and more from the prior distribution, because the data contribute the dominant information,
- the posterior distribution becomes more concentrated, indicating better knowledge of the parameter, less uncertainty, and
- the posterior distribution becomes approximately centered around the MLE, $\hat{\lambda}$.

To be consistent with the notation for random variables, upper case letters would be used for uncertain parameters that have probability distributions. Such notation is not customary in the Bayesian literature, and will not be used here. The reader must judge from context whether the letter θ denotes a particular value, or the uncertain parameter with an associated distribution.

6.2.2.2 Choosing a Prior

The subsections below consider estimation of θ using various possible prior distributions. The simplest prior distribution is discrete. The posterior can be calculated easily, for example, by a spreadsheet. The next simplest prior is called **conjugate**; this prior combines neatly with the likelihood to give a posterior that can be evaluated by simple formulas. Finally, the most general priors are considered; the posterior distribution in such a case can only be found by numerical integration or by random sampling.

The prior distribution should accurately reflect prior knowledge or belief about the unknown parameter. However, quantifying belief is not easy. Raiffa and Schlaifer (1961, Sections 3.3.3-3.3.5) point out that most people can think more easily in terms of percentiles of a distribution than in terms of moments. They also give advice on looking at the situation from many directions, to make sure that the prior belief is internally consistent and has been accurately quantified. Siu and Kelly (1998, Sec. 5.1.4) present seven warnings in connection with developing a prior distribution, which are summarized here:

- Beware of zero values. If the prior says that a value of θ is impossible, no amount of data can overcome this.
- Beware of cognitive biases, caused by the way people tend to think.
- Beware of generating overly narrow prior distributions.
- Ensure that the evidence used to generate the prior distribution is relevant to the estimation problem.
- Be careful when assessing parameters that are not directly observable.
- Beware of conservatism. Realism is the ideal, not conservatism.
- Be careful when using discrete probability distributions.

For a fuller discussion of these points, see Siu and Kelly.

Some priors are chosen to be "noninformative," that is, diffuse enough that they correspond to very little prior information. The **Jeffreys noninformative prior** is often used in this way. If information is available, it is more realistic to build that information into the prior, but sometimes the information is difficult to find and not worth the trouble. In such a case, the Jeffreys noninformative prior can be used. It is one of the priors discussed below.

6.2.2.3 Estimation with a Discrete Prior

When the prior distribution is discrete, the calculations can easily be performed in a spreadsheet. Newcomers to Bayesian estimation are strongly encouraged to work through some examples of the type given here, to develop a sense of how the process works and how the posterior distribution depends on the prior and on the data.

The parameter is assumed to take one of m possible values, $\mathcal{B}_1$, ... , $\mathcal{B}_m$. Let the probability distribution function (p.d.f.) be denoted by f, so $f(\mathcal{B}_i) = \Pr(\mathcal{B}_i)$. This probability quantifies the analyst's prior belief that each of the possible values is the one operating in nature.

Then, some evidence is observed, denoted conceptually by E. Bayes' theorem says that

$$f(\lambda_i|E) = \frac{f(\lambda_i)L(E|\lambda_i)}{\sum_{j=1}^{m} L(E|\lambda_j)f(\lambda_j)} \qquad (6.3)$$

where

$f(\mathcal{B}_i \mid E)$ = the probability of $\mathcal{B}_i$ given evidence E (posterior distribution),

$f(\mathcal{B}_i)$ = the probability of $\mathcal{B}_i$ prior to having evidence E (prior distribution), and

$L(E \mid \mathcal{B}_i)$ = the likelihood function (probability of the evidence given $\mathcal{B}_i$) .

Note that the denominator in Equation 6.3, the total probability of the evidence E, is simply a normalizing constant. Therefore, a more abbreviated form of Bayes' theorem is

$$f(\lambda_i|E_i) \propto f(\lambda_i)L(E|\lambda_i) .$$

This is the form of the theorem that was given in the overview of Section 6.2.2.1.

When the evidence is in the form of x failures generated by a Poisson process over an operational time t, the likelihood function is given by Equation 6.1:

$$L(E|\lambda_i) = e^{-\lambda_i t}\frac{(\lambda_i t)^x}{x!} .$$

The above equations are illustrated here with several prior distributions. For data, they all use the first sample in Example 6.2, 10 events in six years. They all use simple, flat, prior distributions over a moderately wide range, but with different degrees of discreteness. One could argue that this prior is not very informative, but the real reason we choose it is to make the impact of the Bayesian updating process easy to see.

Given the ease of calculation with current computers, a finely discretized prior (say, at 0, 0.01, 0.02,...6.00) would give the most accurate results, and we will provide that calculation in a moment. First, however, let us use a very coarse prior at 0, 0.5, 1.0, ...6.0. With only 13 bins, the reader can perform hand calculations quite easily. The results are given in Table 6.2. The prior is discrete, and is shown in Figure 6.6. The posterior distribution is also discrete, and is shown in Figure 6.7.

Table 6.2 Example 6.2, first sample (10 events in 6 years) with coarse prior.

Event Rate $\mathcal{B}_i$	Prior Probability p_i	Likelihood L_i	p_i x L_i	Posterior Probability Pr($\mathcal{B}_i$|E)	Cumulative Probability E Pr($\mathcal{B}_i$|E)
0.0	0.077	0.00E+0	0.00E+0	0.00E+0	0.00E+0
0.5	0.077	8.10E-4	6.23E-5	2.43E-3	2.43E-3
1.0	0.077	4.13E-2	3.18E-3	1.24E-1	1.26E-1
1.5	0.077	1.19E-1	9.12E-3	3.56E-1	4.82E-1
2.0	0.077	1.05E-1	8.06E-3	3.14E-1	7.96E-1
2.5	0.077	4.86E-2	3.74E-3	1.46E-1	9.42E-1
3.0	0.077	1.50E-2	1.15E-3	4.49E-2	9.87E-1
3.5	0.077	3.49E-3	2.68E-4	1.05E-2	9.98E-1
4.0	0.077	6.60E-4	5.07E-5	1.98E-3	1.00E+0
4.5	0.077	1.07E-4	8.20E-6	3.20E-4	1.00E+0
5.0	0.077	1.52E-5	1.17E-6	4.57E-5	1.00E+0
5.5	0.077	1.97E-6	1.51E-7	5.90E-6	1.00E+0
6.0	0.077	2.34E-7	1.80E-8	7.01E-7	1.00E+0

There is some value in plotting both distributions in the same graph, so they can be compared easily. In such a plot, the vertical bars fall on top of each other and are easily confused. Therefore, we draw the graph by connecting the tops of the bars, in Figure 6.8.

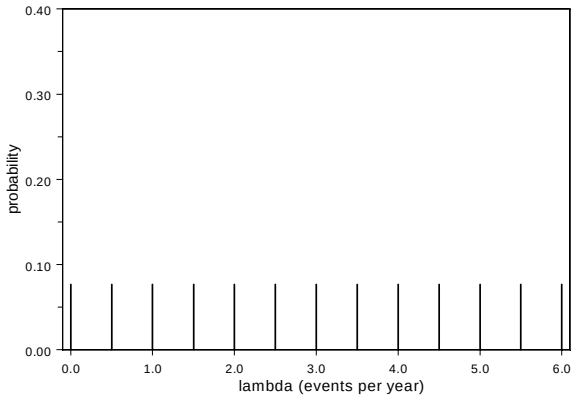


Figure 6.6 Coarse discrete prior distribution for θ.

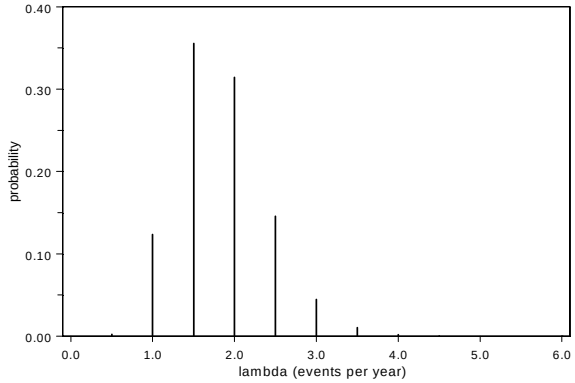


Figure 6.7 Discrete posterior distribution of θ based on 10 events in 6 years.

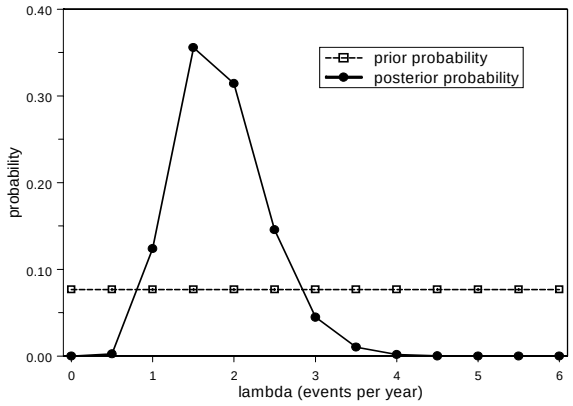


Figure 6.8 Discrete prior and posterior distributions plotted together.

The natural tendency is to think of these curves as densities, but this is not quite correct because they are not normalized to integrate to 1.0. Except for that detail, the curves can be thought of as continuous approximations of the discrete distribution.

Even with such a coarse prior, the evidence is strong and forces the distribution to peak at about θ = 1.5 per year. There is essentially no chance that θ is greater than four or less than 0.5.

If we repeat the calculation with a discrete prior twice as fine (i.e., on the points 0, 0.25, 0.50, 0.75,...6.00), the prior now has 25 bins and the results are much more smooth, as shown in Figure 6.9. These results are quite smooth, and of course follow the previous results.

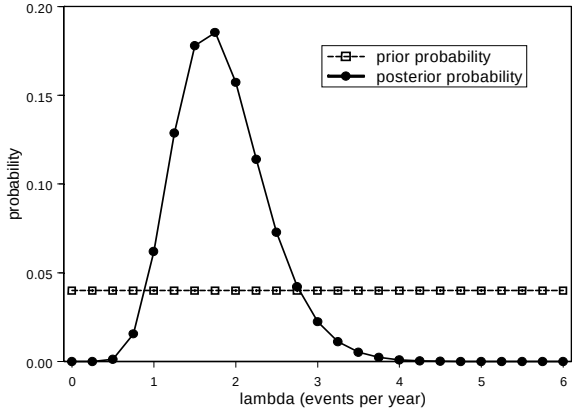


Figure 6.9 Discrete prior and posterior distributions for 10 events in 6 years, with finely discretized prior.

Finally, let us repeat the calculation for a discrete flat prior on the points 0, 0.05, 0.10, 0.15,...6.00, i.e., a 121-point grid. This time, the results, shown in Figure 6.10, are detailed enough to closely approximate a smooth, continuous distribution.

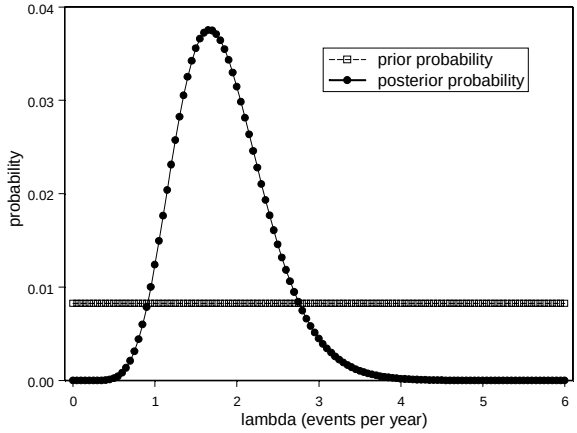


Figure 6.10 Discrete prior and posterior distributions for 10 events in six years, with very finely discretized prior.

The spreadsheet calculation is identical with Table 6.2, except for having 121 bins rather than 13. Some values are summarized in Table 6.3. These Bayesian results are also compared with the frequentist estimates obtained from the first sample, shown earlier in Figure 6.4. The Bayesian posterior distribution has a mode where the probability function is maximized, and a mean equal to $E\theta$, $f(\theta)$.

Table 6.3. Comparison of Bayesian and frequentist estimates for the data in Example 6.2.

Quantity	Bayes, flat prior	Frequentist, Figure 6.4
Point estimate (Bayes mode, Bayes mean, frequentist MLE)	1.65 1.833	1.73
Lower end of interval (Bayes 5th percentile, lower confidence limit)	1.00	0.95
Upper end of interval (Bayes 95th percentile, upper confidence limit)	2.80	2.85

These values are compared to the frequentist point estimate, the MLE. The Bayesian 5th and 95th percentiles form a **Bayes credible interval**, which is compared with the frequentist confidence limits shown at the top of Figure 6.4. The Bayes 90% interval, based on a flat, essentially noninformative prior, is slightly more narrow than the frequentist 90% confidence interval.

This concludes the examples for this section. However, we suggest that the reader make up a data set for examining the way the posterior distribution responds to growing evidence. For example, try beginning with zero failures in year one; then adding two failures in year two; then zero failures in year three; etc. Also try a case that does not agree with the prior; for example five failures in year one; then seven more in year two; then six in year three. Such examples are given for p in Section 6.3.2.1, but they are most valuable to someone who constructs them and works them out, instead of merely reading about them.

6.2.2.4 Estimation with a Conjugate Prior

We now turn from discrete to continuous prior distributions. We begin with a very convenient family of distributions: the conjugate priors.

6.2.2.4.1 Definitions

The conjugate family of prior distributions for Poisson data is the family of gamma distributions. Two parameterizations of gamma distributions are given in Appendix A.7.6. For Bayesian estimation, the following parameterization is the more convenient one:

$$f(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda\beta} \quad . \tag{6.4}$$

Here, θ has units 1/time and β has units of time, so the product $\theta\beta$ is unitless. For example, if θ is the frequency of events per critical-year, β has units of critical-years. The parameter β is a kind of **scale parameter**. That is, β corresponds to the scale of θ. If we convert θ from events per hour to events per year by multiplying it by 8760, we correspondingly divide β by 8760, converting it from hours to years. The other parameter, α, is unitless, and is called the **shape parameter**. The gamma function, $\Gamma(\alpha)$, is a standard mathematical function, defined in Appendix A.7.6. If α is a positive integer, $\Gamma(\alpha)$ equals $(\alpha-1)!$

Let θ have a gamma uncertainty distribution. In the present parameterization, the mean of the gamma distribution, also written as the expected value $E(\theta)$, is α/β, and the variance, var(θ), is α/β^2. Note that the units are correct, units 1/time for the mean and 1/time^2 for the variance.

6.2.2.4.2 Update Formulas

As stated earlier and in Appendix B.5.1, the posterior distribution is related to the prior distribution by

$$f_{\text{post}}(\lambda) \propto \Pr(X = x \mid \lambda) f_{\text{prior}}(\lambda) \tag{6.5}$$

This is the continuous analogue of Equation 6.3. The probability of the data is also called the **likelihood**, in which case it is considered as a function of the parameter θ for a given x. For Poisson data, it is given by Equation 6.1. The symbol $\propto$ denotes "is proportional to." Probability density functions generally have normalizing constants in front to make them integrate to 1.0. These constants can be complicated, but using proportionality instead of equality allows us to neglect the normalizing constants. Stripped of all the normalizing constants, the gamma p.d.f. is

$$f(\lambda) \propto \lambda^{\alpha-1} e^{-\lambda\beta} \quad .$$

The gamma distribution and the Poisson likelihood combine in a beautifully convenient way:

$$f_{\text{post}}(\lambda) \propto e^{-\lambda t}\frac{(\lambda t)^x}{x!}\lambda^{\alpha-1}e^{-\lambda\beta}$$
$$\propto \lambda^{(x+\alpha)-1}e^{-\lambda(t+\beta)}$$

In the final expression, everything that does not involve θ has been absorbed into the proportionality constant. This result is "beautifully convenient," because the posterior distribution of θ is again a gamma distribution. This is the meaning of **conjugate**: if the prior distribution is a member of the family (in this case, the gamma family), the posterior distribution is a member of the same family. The update formulas are:

$$\alpha_{\text{post}} = x + \alpha_{\text{prior}}$$
$$\beta_{\text{post}} = t + \beta_{\text{prior}}$$

This leads to an intuitive interpretation of the prior parameters: a gamma(α_{prior}, β_{prior}) distribution is equivalent, at least intuitively, to having seen α_{prior} events in β_{prior} time units, prior to taking the current data.

Figure 6.5 was constructed in this way. The prior distribution was gamma(0.2, 10,000). Therefore, the posterior distributions were gamma(1.2, 20,000), gamma(10.2, 110,000), and gamma(50.2, 510,000).

When using these update formulas, be sure that t and β_{prior} have the same units. If one is expressed in hours and one in years, one of the two numbers must be converted before the two are added.

The moments of the gamma distribution were mentioned previously. The posterior mean is $\alpha_{\text{post}}/\beta_{\text{post}}$ and the posterior variance is $\alpha_{\text{post}}/(\beta_{\text{post}})^2$.

The percentiles of the gamma distribution are given by many software packages. If you use such software, be careful to check that it is using the same parameterization that is used here! Here are three ways to get the correct answer. (1) If the software uses the other parameterization, fool it by inverting your value of β. Then check to make sure that the numbers appear reasonable. (2) A safe method is to have the software find the percentiles of the gamma(α_{post}, 1) distribution. Then manually divide these percentiles by β_{post}. This ensures that the scale parameter is treated correctly. (3) As a final alternative, the percentiles of the gamma distribution can be found from a tabulation of the chi-squared distribution, possibly interpolating the table. To do this, denote the (100p)th percentile of the posterior distribution by θ_p. For example, denote the 95th percentile by $\theta_{0.95}$. The (100p)th percentile is given by:

$$\theta_p = \chi^2_p(2\alpha_{\text{post}})/(2\beta_{\text{post}})$$

where, as elsewhere, $\chi^2_p(d)$ is the pth quantile, or (100p)th percentile, of a chi-squared distribution with d degrees of freedom. Note the presence of 2 in the numerator and denominator when the chi-squared distribution is used.

The next section contains examples that use these update formulas with several priors.

6.2.2.5 Possible Conjugate Priors

6.2.2.5.1 Informative Priors

The prior distribution must come from sources other than the current data. It might be tempting to use the data when constructing the prior distribution, but that temptation must be resisted. Prior distributions are named "prior" for a reason: they reflect information that does not come from the current data. Ideally, generic data provide the basis for prior belief. Generic data sources are given in Section 4.2.

Consider again Example 6.1, involving initiating events with loss of heat sink. With no special knowledge about the plant, prior belief about the plant is reasonably based on the overall industry performance, so we use the generic industry distribution as the prior. Poloski et al. (1999a) examined initiating-event data from the nuclear power industry over nine years. For PWRs, and initiating events involving loss of heat sink, they determined that the variability of θ across the industry can be described by a gamma distribution with shape parameter = 1.53, and scale parameter = 10.63 reactor-critical-years. Regrettably, Table G-1 of the report gives only a mean and a 90% interval, not the distribution and its parameters. The distribution given here is taken from the unpublished work that formed the basis of the report. The distribution is a gamma distribution, so the update formulas given above can be used in the hypothetical example of this section. The prior distribution is shown in Figure 6.11.

Now, consider updating this prior with the data from Example 6.1. To make the units consistent, convert the 42800 reactor-critical-hours in the example to 42800/8760 = 4.89 reactor-critical-years. The update formula yields:

$\alpha_{\text{post}} = x + \alpha_{\text{prior}} = 1 + 1.53 = 2.53$

$\beta_{\text{post}} = t + \beta_{\text{prior}} = 4.89 + 10.63 = 15.52$ reactor-critical-years.

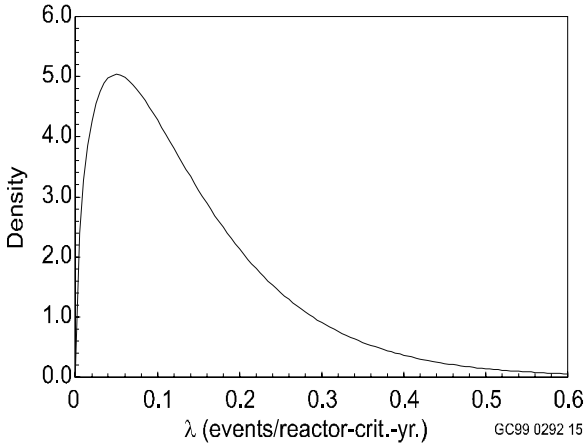


Figure 6.11 Prior density for θ, gamma(1.53, 10.63).

The mean, μ_{post}/ξ_{post}, is 0.163 events per reactor-critical-year, the variance is 0.0105 (per reactor-critical-year squared), and the standard deviation is the square root of the variance, 0.102 per reactor-critical-year.

A 90% credible interval is the interval from the 5th to the 95th percentiles of the posterior distribution. A software package finds the two percentiles of a gamma(2.53, 1.0) distribution to be 0.5867 and 5.5817. Division by ξ_{post} yields the two percentiles of the posterior distribution: 0.038 and 0.36. Alternatively, one may interpolate Table C.2 of Appendix C to find the percentiles of a chi-squared distribution with 5.06 degrees of freedom, and divide these percentiles by $2\xi_{post}$. Linear interpolation gives answers that agree to three significant digits with the exact answers, but if the degrees of freedom had not been so close to an integer, the linear interpolation might have introduced a small inaccuracy.

The interpretation of the above numbers is the following. The best belief is that θ is around 0.16, although it could easily be somewhat larger or smaller. Values as small as 0.038 or as large as 0.36 are possible, but are approaching the limits of credibility.

Two graphical ways of presenting this information are given below. Figure 6.12 shows the posterior density. The areas to the left of the 5th percentile and to the right of the 95th percentile are shaded. The 90% credible interval is the interval in the middle, with probability 90%. Figure 6.13 shows the same information using the cumulative distribution. The 5th and 95th percentiles are the values of θ where the cumulative distribution is 0.05 and 0.95, respectively. These percentiles are the same values, as shown in the plot of the density.

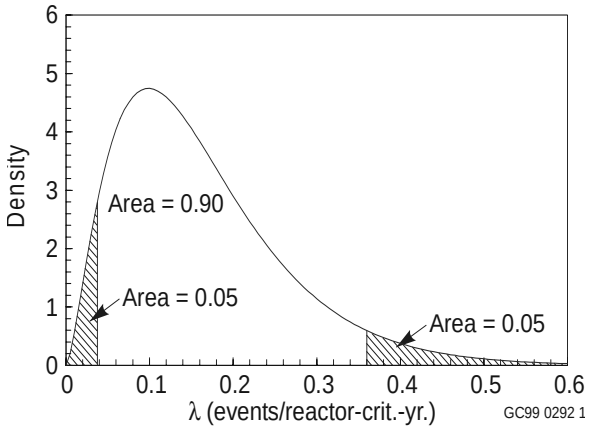


Figure 6.12 Posterior density of θ, gamma(2.53, 15.52), for Example 6.1 with industry prior. The 5th and 95th percentiles are shown.

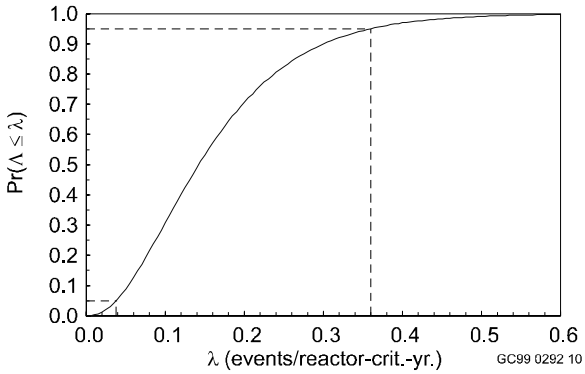


Figure 6.13 Posterior cumulative distribution of θ for Example 6.1 with industry prior. The 5th and 95th percentiles are shown.

For PRA applications, however, the right tail is typically of concern for risk, corresponding to high initiating event frequency (or, in other sections of this chapter, high probability of failure on demand, high unavailability, or long time to recovery). The interval given above holds the error probability for the right tail equal to 0.05. This number is customary in much statistical practice, and has therefore been used in many studies for the NRC. The lower end of the interval, on the other hand, is not of great safety concern. It is easy to calculate, however. Therefore, the above 90% interval, corresponding to 5% posterior probability in each tail, is commonly presented in PRA studies.

Actually, however, the interval presents only a portion of the information in the posterior distribution, two summary numbers. The full distribution is used in a PRA.

6.2.2.5.2 Noninformative Prior

The **Jeffreys noninformative prior** is intended to convey little prior belief or information, thus allowing the data to speak for themselves. This is useful when no informed consensus exists about the true value of the unknown parameter. It is also useful when the prior distribution may be challenged by people with various agendas. Some authors use the term **reference prior** instead of "noninformative prior," suggesting that the prior is a standard default, a prior that allows consistency and comparability from one study to another.

With Poisson data, the Jeffreys noninformative prior is obtained if the shape parameter of a gamma distribution is taken to be $\alpha = \frac{1}{2}$ and the parameter β is taken to be zero. (See, for example, Box and Tiao 1973.) Ignoring the normalizing constant at the front of Equation 6.4 yields a function that is proportional to $\theta^{-\frac{1}{2}}$, shown in Figure 6.14. Although this function is interpreted as a density function, it is an **improper distribution** because its integral from 0 to ∞ is infinite.

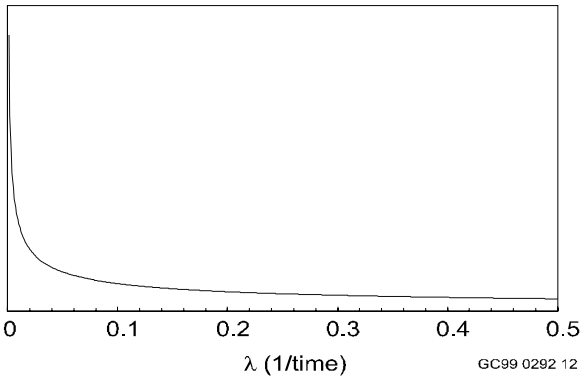


Figure 6.14 Jeffreys noninformative prior distribution for an event frequency.

It is not intuitive that this prior is "noninformative." Simple intuition might expect a uniform distribution instead. To better educate the intuition, suppose we had some median prior value m; that is, the prior distribution of θ satisfies $\Pr(\theta < m) = \Pr(\theta > m)$. This can be rewritten as

$$\Pr(0 < \theta < m) = \Pr(m < \theta < \infty) .$$

The interval from 0 to m is shorter than the interval from m to ∞. Therefore, the prior density should be larger to the left of m than to the right. The density shown in Figure 6.14 has this property. (We ignore the fact that the density in Figure 6.14 is improper.)

Further arguments for the prior are too complicated to give here. For a fuller explanation of noninformative priors, see Appendix B.5.3.1 and the references cited there.

Suppose that the data consist of x events in time t. Formal application of the update formulas yields

$$\alpha_{post} = x + \tfrac{1}{2}$$
$$\beta_{post} = t + 0 .$$

That is, the Bayes posterior distribution for θ is gamma($x + \frac{1}{2}$, t).

It is interesting to compare the interval using the Jeffreys prior with the corresponding confidence interval. The 90% posterior credible interval is

$$\theta_{0.05} = \chi^2_{0.05}(2x + 1)/2t$$
$$\theta_{0.95} = \chi^2_{0.95}(2x + 1)/2t.$$

These may be compared with the 90% confidence interval:

$$\theta_{conf,\ 0.05} = \chi^2_{0.05}(2x)/2t$$
$$\theta_{conf,\ 0.95} = \chi^2_{0.95}(2x + 2)/2t.$$

The confidence intervals differ from the Bayes credible intervals only in the degrees of freedom, and there only slightly. This is the primary sense in which the Jeffreys prior is "noninformative." The lower and upper *confidence* limits have degrees of freedom $2x$ and $2x + 2$, respectively. The two *Bayesian* limits each use the average, $2x + 1$. The confidence interval is wider than the Jeffreys credible interval, a reflection of the conservatism of confidence limits with discrete data. However the similarity between the confidence limits and the Jeffreys limits shows that the result using the Jeffreys prior will resemble the result using frequentist methods, that is, using no prior information at all.

Consider again Example 6.1, with one event in 4.89 critical-years, and use the Jeffreys noninformative prior. The resulting posterior distribution has

$$\alpha_{post} = 1.5$$
$$\beta_{post} = 4.89 \text{ critical-years} .$$

The mean of this distribution is 1.5/4.89 = 0.31 events per critical-year. A 90% Bayes credible interval can be obtained from a chi-squared table without any need for interpolation, because the degrees of freedom parameter is 3, an integer. The 5th and 95th percentiles of the chi-squared distribution are 0.352 and 7.815. Division by 2×4.89 yields the percentiles of the posterior distribution, 0.036 and 0.80.

This posterior distribution has a larger mean and larger percentiles than the posterior distribution in Section 6.2.2.5.1. The data set is the same, but the different prior distribution results in a different posterior distribution. The results will be compared in Section 6.2.2.5.4.

6.2.2.5.3 Constrained Noninformative Prior

This prior is a compromise between an informative prior and the Jeffreys noninformative prior. The mean of the constrained noninformative prior uses prior belief, but the dispersion is defined to correspond to little information. These priors are described by Atwood (1996), and by references given there. Constrained noninformative priors have not been widely used, but they are mentioned here for the sake of completeness.

For Poisson data, the constrained noninformative prior is a gamma distribution, with the mean given by prior belief and the shape parameter = ½. That is:

$\alpha_{prior} = ½$
β_{prior} satisfies $\alpha_{prior}/\beta_{prior}$ = prior mean .

To illustrate the computations, consider again the Example 6.1, with one event in 4.89 reactor-critical-years. Suppose we knew that in the industry overall such events occur with an average frequency of 0.144 events per reactor-critical-year. (This is consistent with the informative prior given above in Section 6.2.2.5.1.) Suppose further that we were unable or unwilling to make any statement about the dispersion around this mean — the full information used to construct the informative prior was not available, or the plant under consideration was atypical in some way, so that a more diffuse prior was appropriate.

The constrained noninformative prior with mean 0.144 has α_{prior} = ½ and β_{prior} = 3.47 critical-years. The resulting posterior distribution has

$\alpha_{post} = x + ½ = 1.5$
$\beta_{post} = t + 3.47 = 8.36$

The mean is 0.18 events per critical-year, and the 90% credible interval is (0.021, 0.47). This notation means the interval from 0.021 to 0.47.

6.2.2.5.4 Example Comparisons Using Above Priors

In general, the following statements can be made:

- The Jeffreys noninformative prior results in a posterior credible interval that is numerically similar to a confidence interval, but slightly shorter.
- If the prior mean exists, the posterior mean is between the prior mean and the MLE.
- If two prior distributions have the same mean, the more concentrated (less diffuse) prior distribution will yield the more concentrated posterior distribution, and will pull the posterior mean closer to the prior mean.

These statements are now illustrated by example. The estimates found in the above sections for Example 6.2 and the various priors are compared in Table 6.4 and in Figure 6.15.

Table 6.4 Comparison of estimates with 1 event in 4.89 reactor-critical-years.

Method	Prior mean	Posterior parameters	Point estimate (MLE or posterior mean)	90% interval (confidence interval or posterior credible interval)
Frequentist	NA	NA	0.20	(0.010, 0.97)
Bayes with Jeffreys noninformative prior, gamma(0.5, 0)	undefined	α = 1.5 β = 4.89	0.31	(0.036, 0.80)
Bayes with (informative) industry prior, gamma(1.53, 10.63)	0.144	α = 2.53 β = 15.52	0.16	(0.038, 0.36)
Bayes with constrained noninformative prior, gamma(0.5, 3.47)	0.144	α = 1.5 β = 8.36	0.18	(0.021, 0.47)

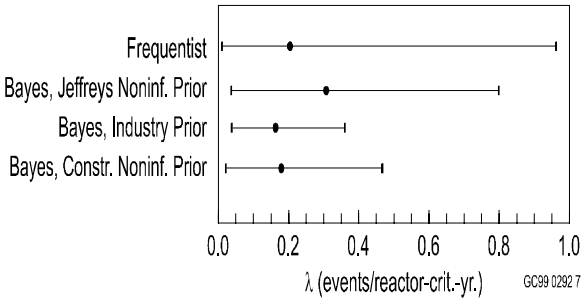


Figure 6.15 Comparison of four point estimates and interval estimates for θ.

In Table 6.4 and in Figure 6.15, the Jeffreys prior and the frequentist approach are listed next to each other because they give numerically similar results. The Jeffreys prior yields a posterior credible interval that resembles the frequentist confidence interval. It is a little shorter, but it is neither to the right nor to the left. This agrees with the earlier discussion of the Jeffreys prior.

In each Bayesian case, the posterior mean falls between the prior mean (if defined) and the MLE, 0.20. The prior distribution has more influence when the prior distribution is more tightly concentrated around the mean. The concentration is measured by the shape parameter α_{prior}, because $1/\alpha$ equals the relative variance (= variance/mean^2). Therefore the larger α, the smaller the relative variance. The industry prior and the constrained noninformative prior have the same mean, but the industry prior has the larger α, that is, the smaller variance. As a consequence, in both cases the posterior mean is between the MLE, 0.204, and the prior mean, 0.144, but the posterior mean based on the industry prior is closer to 0.144, because that prior has a smaller variance. Because the prior mean is smaller than the MLE, the bottom two lines give smaller posterior estimates than do the top two lines. Also, the prior distribution with the most information (largest α) yields the most concentrated posterior distribution, and the shortest 90% interval.

In some situations, no conjugate prior is satisfactory. For example, a gamma distribution is very unrealistic if the shape parameter is very small. As a rule of thumb, the lower percentiles of the distribution are unrealistic if α is much smaller than 0.5. Such a posterior distribution arises with Poisson data when the prior distribution is very skewed (α very small) and the data contain zero events. Then, the posterior distribution also is very skewed, and the posterior 5th percentile may be many orders of magnitude below the posterior mean. The subject-matter experts must look at the percentiles and decide if they are believable. If not, a more appropriate prior should be chosen. It will not be conjugate. This is the subject of the next subsection.

6.2.2.6 Estimation with a Continuous Nonconjugate Prior

Discrete priors and conjugate priors were updated above with simple formulas. What remains are the continuous nonconjugate priors. Any continuous distribution defined on the allowed range of θ can, in principle, be used as a prior. The resulting posterior distribution is a continuous distribution, with no simple form. (Because the posterior distribution does not have a simple analytical form, it cannot be entered directly as an input to most PRA codes. Instead, a discrete approximation of the posterior distribution must usually be used.)

Three approaches for obtaining the posterior are given here. Some examples will be worked out in Section 6.2.2.7.

6.2.2.6.1 Direct Numerical Integration

If software is available for performing numerical integration, the following approach can be used. Find the form of the posterior distribution, using Equation 6.5. Suppose, for example, that the prior distribution for θ is lognormal, with μ and σ^2 denoting the mean and variance of the normal distribution of $\ln\theta$. As stated in Appendix A.7.3, the lognormal density is proportional to

$$f_{LN}(\lambda) \propto \frac{1}{\lambda} e^{-\frac{1}{2}\left(\frac{\ln\lambda-\mu}{\sigma}\right)^2} .$$

Substitute this and Equation 6.1 into Equation 6.5, to obtain the form of the posterior density:

$$Cf_{post}(\lambda) = e^{-\lambda t}\lambda^x \frac{1}{\lambda} e^{-\frac{1}{2}\left(\frac{\ln\lambda-\mu}{\sigma}\right)^2} .$$

All terms that do not involve θ have been absorbed into the normalizing constant, C. The normalizing constant can be evaluated by numerically integrating Cf_{post} from 0 to 4, that is, integrate the right hand side of the equation. Unless x is unrealistically large, the function does not need to be integrated in practice out beyond, say, $\ln\theta = \mu + 5\sigma$. C equals the integral of Cf_{post}, because the integral of f_{post} must equal 1. Once C has been evaluated, the mean and percentiles of f_{post} can be found numerically.

Numerical integration, using a technique such as the trapezoidal rule or Simpson's rule, can be programmed easily, even in a spreadsheet. The ideas are found in some calculus texts, and in books on numerical methods such as Press et al. (1992).

6.2.2.6.2 Simple Random Sampling

A second approach, which does not directly involve numerical integration, is to generate a large random sample from the posterior distribution, and use the sample to approximate the properties of the distribution. Some people think of this as numerical integration via random sampling. Surprisingly, the random sample can be generated without explicitly finding the form of the posterior distribution, as explained by Smith and Gelfand (1992).

The algorithm, called the **rejection method** for sampling from a distribution, is given here in its general form, and applied immediately to sampling from the posterior distribution. In general, suppose that it is possible to sample some parameter θ from a continuous distribution g, but that sampling from a different distribution f is desired. Suppose also that a positive constant M can be found such that $f(\theta)/g(\theta) \leq M$ for all θ. The algorithm is:

(1) Generate θ from $g(\theta)$;
(2) Generate u from a uniform distribution, $0 \leq u \leq 1$; and
(3) If $u \leq f(\theta)/[Mg(\theta)]$ accept θ in the sample. Otherwise discard it.

Repeat Steps (1) through (3) until enough values of θ have been accepted to form a sample of the desired size.

This algorithm is the basis for many random-number generation routines in software packages. It is applied below to the generation of a sample from the posterior distribution for θ. The equations are worked out here, and the algorithm for the posterior distribution is restated at the end.

Let f be the posterior density and let g be the prior density. Then Equation 6.5 states that the ratio $f(\theta)/g(\theta)$ is proportional to the likelihood, which is maximized, by definition, when θ equals the maximum likelihood estimate, x/t. That is, the ratio of interest is:

$$f(\theta)/g(\theta) = Ce^{-\theta t}(\theta t)^x$$

for some constant C. This is maximized when θ equals x/t. Therefore, define $M = \max[f(\theta)/g(\theta)] = Ce^{-x}x^x$. The condition in Step (3) above is equivalent to:

$$u \leq [Ce^{-\theta t}(\theta t)^x] / [Ce^{-x}x^x] = [e^{-\theta t}(\theta t)^x] / [e^{-x}x^x] .$$

The constant cancels in the numerator and denominator, so we do not need to evaluate it! It would have been possible to work with $m = M/C$, and the calculations would have been simpler. This rewritten form of the algorithm, for Poisson data, is given here.

If $x > 0$, define $m = e^{1-x}x^x$. If $x = 0$, define $m = 1$.

The steps of the algorithm are:

(1) Generate a random θ from the prior distribution;
(2) Generate u from a uniform distribution, $0 \leq u \leq 1$; and
(3) If $u \leq e^{-\theta t}(\theta t)^x/m$, accept θ in the sample. Otherwise discard θ.

Repeat Steps (1) through (3) until a sample of the desired size is found.

Intuitively, this algorithm generates possible values of θ from the prior distribution, and discards most of those that are not very consistent with the data. The result is a sample from the posterior distribution.

6.2.2.6.3 More Complicated Random Sampling

All-purpose Bayesian update programs can be used for the present simple problem. For example, the program BUGS[1] (Bayesian inference Using Gibbs Sampling) performs **Markov chain Monte Carlo (MCMC)** sampling. This package is intended for complicated settings, such as those described in Chapters 7 and 8. Using it here is like using the proverbial cannon to kill a mosquito. Nevertheless, the program is free, and very flexible, and can be used here. It is available for download at

http://www.mrc-bsu.cam.ac.uk/bugs/

and is described more fully in Sections 7.2.3 and 8.3.3.3 of this handbook. An example is given below.

---

[1] Mention of specific products and/or manufacturers in this document implies neither endorsement or preference, nor disapproval by the U.S. Government or any of its agencies of the use of a specific product for any purpose.

6.2.2.7 Examples Involving Nonconjugate Priors

These techniques will be illustrated with Example 6.3, from Appendix J-4 of Poloski et al. (1999a).

Example 6.3 Small-break LOCAs.

> No small-break loss-of-coolant accidents (SBLOCAs) have occurred in 2102 reactor-calendar-years at U.S. nuclear power plants. The WASH-1400 (NRC 1975) distribution for the frequency of this event was lognormal with median 1E!3 and error factor 10.

6.2.2.7.1 Example with Lognormal Prior

Poloski et al. (1999a) use the WASH-1400 distribution as a prior, and update it with the 2102 years of data.

The resulting posterior distribution was sampled 100,000 times using the method described in Section 6.2.2.6.2 above, and the mean was found. Then, the values were arranged in increasing order, and the percentiles of the sample were found. This process took less than 15 seconds in 1999 on a 166 MHz computer. Based on the mean and percentiles of the sample, the mean of the posterior distribution is 3.5E!4, and the 90% posterior credible interval is (4.5E!5, 9.8E!4).

To illustrate the method of Section 6.2.2.6.3, the distribution was also sampled using BUGS. Figure 6.16 shows the script used for running BUGS.

```
model
{
  mu <- lambda*rxyrs
  x ~ dpois(mu)
  lambda ~ dlnorm(-6.908, 0.5104)
}
list(rxyrs=2102, x=0)
```

Figure 6.16 Script for analyzing Example 6.3 using BUGS.

The section in curly brackets defines the model. Note that <-, intended to look like a left-pointing arrow, is used to define quantities in terms of other quantities, and ~ is used to generate a random quantity from a distribution. The names of distributions begin with the letter *d*. Thus, X is a Poisson random variable with mean λ, with $\lambda = \theta \times$ rxyrs. The prior distribution of θ is lognormal. The parameters given in the script arise as follows. BUGS parameterizes the normal in terms of the mean and inverse of the variance, for reasons explained in Section 6.6.1.2.1. It parameterizes the lognormal distribution using the parameters of the underlying normal. It is shown below that a lognormal with median 1E!3 and error factor 10 corresponds to an underlying normal with mean !6.980 and standard deviation 1.3997. Therefore, the inverse of the variance is $1/1.3997^2 = 0.5104$.

The line beginning "list" defines the data, 0 events is 2102 reactor years. BUGS also requires an initial value for θ, but generated it randomly.

When BUGS generated 100,000 samples, the mean, 5th percentile, and 95th percentile of θ were 3.5E!4, 4.5E!5, and 9.8E!4, just as found above.

6.2.2.7.2 Example with "Moment-Matching" Conjugate Prior

Conjugate priors have appeal: Some people find algebraic formulas tidier and more convenient than brute-force computer calculations. Also, when a PRA program requests a distribution for a parameter, it is usually easier to enter a distributional form and a couple of parameters than to enter a simulated distribution.

Therefore, a nonconjugate prior is sometimes replaced by a conjugate prior having the same mean and variance. This method is carried out here with the above example.

Begin by finding the gamma prior with the same moments as the above lognormal prior. As explained in Appendix A.7.3, the median, error factor, and moments of the lognormal distribution are related to λ and ζ of the underlying normal distribution of ln7 as follows:

$$\text{median}(\theta) = \exp(\lambda)$$
$$\text{EF}(\theta) = \exp(1.645\zeta)$$
$$\text{mean}(\theta) = \exp(\lambda + \zeta^2/2)$$
$$\text{var}(\theta) = [\text{median}(\theta)]^2 \exp(\zeta^2)[\exp(\zeta^2) - 1] .$$

The lognormal prior has median 1.0E!3, and error factor 10. Solving the first two equations yields

$$\lambda = -6.907755$$
$$\zeta = 1.399748 .$$

Substituting these values into the second two equations yields

$$\text{mean}(\theta) = 2.6635\text{E}{-3}$$
$$\text{var}(\theta) = 4.3235\text{E}{-5} .$$

Now the gamma distribution must be found with this mean and this variance. The formulas for the moments of a gamma distribution were given in Section 6.2.2.4.1 and in Appendix A.7.6:

mean = α/β
variance = α/β^2 .

Therefore,

α = mean^2/variance = 0.164
β = mean/variance = 61.6 reactor-years.

Warning flags should go up, because α is considerably smaller than 0.5. Nevertheless, we carry out the example using this gamma distribution as the prior. The update formulas yield:

α_{post} = 0 + 0.164 = 0.164
β_{post} = 2102 + 61.6 = 2164 reactor-years .

The posterior mean is 7.6E!5, and a 90% credible interval is (3.4E!12, 4.1E!4), all with units events per reactor-year.

6.2.2.7.3 Comparison of Example Analyses

The two posterior distributions do not agree closely, as will be discussed below. If the shape parameter α of the gamma prior had been larger, the two prior distributions would have had more similar percentiles, and the two posterior distributions likewise would have agreed better. As it is, however, the two analyses are summarized in Table 6.5.

Table 6.5 Posterior distributions from two analyses.

Prior	Mean	90% Interval
Lognormal	3.5E!4	(4.5E!5, 9.8E!4)
Gamma	7.6E!5	(3.4E!12, 4.1E!4)

The most notable difference between the two posterior distributions is in the lower endpoints, the 5th percentiles, which differ by many orders of magnitude. This is explained, to some extent, by graphical comparisons. Figures 6.17 and 6.18 show the prior cumulative distributions. When plotted on an ordinary scale in Figure 6.17, the two prior distributions look fairly similar, although the gamma distribution seems to put more probability near zero. The differences become much more obvious when the two prior distributions are plotted on a logarithmic scale in Figure 6.18. These differences between the two prior distributions are present in spite of the fact that the two priors have equal means and equal variances.

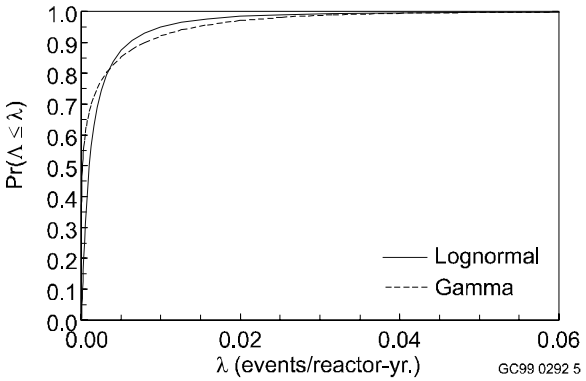


Figure 6.17 Two prior distributions having the same means and variances.

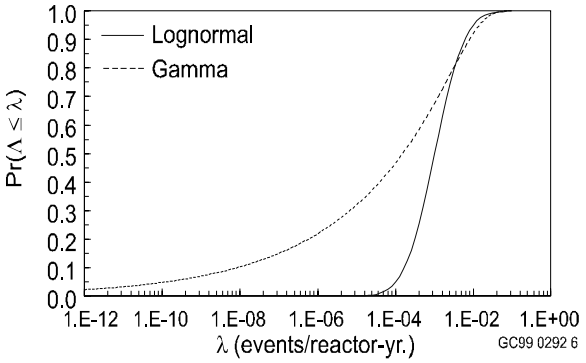


Figure 6.18 The same prior distributions as in the previous figure, with β plotted on a logarithmic scale.

The two resulting posterior distributions are also quite different in the lower tail, as shown in Figure 6.19, and this difference is especially clear when the distributions are plotted on a log scale, as shown in Figure 6.20.

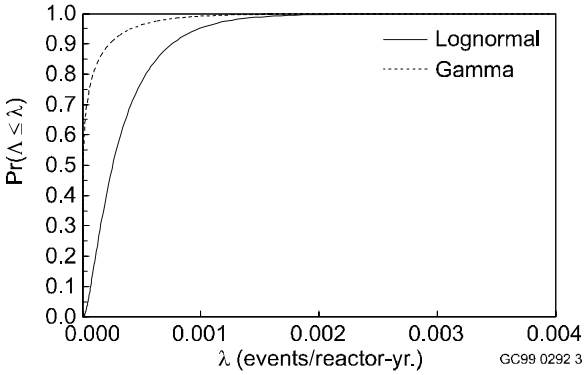


Figure 6.19 Two posterior distributions, from priors in previous figures.

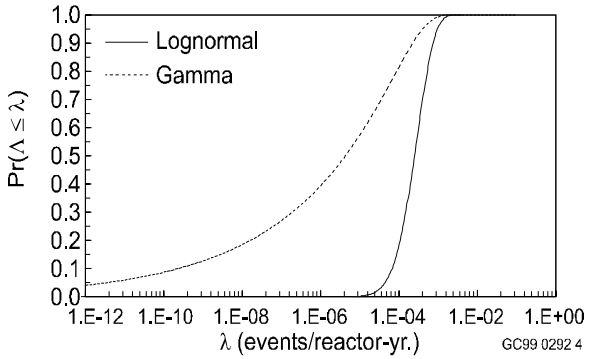


Figure 6.20 The same posterior distributions as in the previous figure, with β plotted on logarithmic scale.

Incidentally, these illustrations use cumulative distributions instead of densities, for an important reason. Cumulative distributions simply show probabilities, and so can be plotted with the horizontal scale either linear or logarithmic. Alternatively, the density of $\ln(\beta)$ could be plotted against $\ln(\beta)$, but take care to calculate the density of $\ln(\beta)$ correctly, as explained in Appendix A.4.7.

6.2.2.8 Analysis with Fixed Count and Random Time

Sometimes it is useful to consider a fixed number of events in a random time, a **waiting time**. For example, if the event frequency is believed to change over time, only the most recent history may represent current behavior. In such a situation, one might decide to use only the most recent few events, such as $x = 3$, and to treat the corresponding time t as random. Here t is the time measured backwards from the present to the xth event in the past. Earlier events could be used to construct a prior distribution, but the dispersion of the prior distribution should be set large because the earlier events are not considered fully relevant to the present.

The above data consist of x exponential(β) durations. The analysis techniques are given in Section 6.6.1.2.2. It turns out that Bayesian analysis with an informative prior is exactly the same whether the data are regarded as coming from a Poisson count x in fixed time t or a sum t of x exponential durations. The two likelihoods are proportional to each other, and the posterior distributions are identical.

6.2.3 Model Validation

Model validation should go hand in hand with parameter estimation. Philosophically, it would seem natural first to confirm the form of the model, and second to estimate the parameters of that model. However, typically one can perform goodness-of-fit tests, and other validations of a model, only after the model has been fully specified, that is, only after the form of the model has been assumed *and* the corresponding parameters have been estimated. Because parameter-estimation is built into most model-validation procedures, it was presented first.

It is usually wise not to stop the analysis with just estimating the parameters. Foolish results have been presented by analysts who estimated the parameters but did not thoroughly check that the assumptions of the model were correct. This section presents ways to check the model assumptions.

That being said, there is more in this section than will be needed on any one analysis. Often, a simple plot is sufficient to show that the model appears adequate. When the data are very sparse, perhaps not even that is needed, because the data set is too small to invalidate any model; in such a case, the simplest model is normally accepted. The methods here are offered for possible use, and the analyst should select the appropriate ones.

The Poisson process was introduced in Section 2.2.2. The three assumptions were listed there: constant event occurrence rate, no simultaneous events, and independent time periods. These assumptions are considered here. Much of the following material is taken from an INEEL report by Engelhardt (1994).

The assumption of constant rate is considered in the next two sections, first, where the alternative possibility is that different data sources may have different values of β, but in no particular order, and then, where the alternative possibility is that a time trend exists. Both graphical methods and formal statistical hypothesis tests are given for addressing the issues. The assumption of no exactly simultaneous events is then discussed from the viewpoint of examining the data for common-cause events. Finally, the assumption of independent time intervals is considered, and some statistical tests of the assumption are given.

When Bayesian methods are used, one must also examine whether the data and the prior distribution are consistent. It makes little sense to update a prior with data, if the data make it clear that the prior belief was incorrect. That topic constitutes the final subsection of the present section.

6.2.3.1 Poolability of Data Subsets

Assumption 1 in Section 2.2.2 implies that there is one rate β for the entire process. The correctness of such an assumption can be investigated by analyzing subsets of

the data and comparing the estimates of θ for the various subsets.

Example 2.2 described LOSP events during shutdown. For this section, consider a portion of that example. The entire data set could be used, but to keep the example from being too cumbersome we arbitrarily restrict it to five plants at three sites, all located in one state.

An obvious question concerns the possibility of different rates for different plants. A general term used in this handbook will be **data subsets**. In Example 6.4, five subsets are shown, corresponding to plants. In other examples, the subsets could correspond to years, or systems, or any other way of splitting the data. For initiating events, each subset corresponds to one **cell** in the table, with an event count and an exposure time.

Example 6.4 Shutdown LOSP events at five plants, 1980-96.

During 1980-1996, five plants experienced eight LOSP events while in shutdown. These were events from plant-centered causes rather than external causes. The data are given here.

Plant code	Events	Plant shutdown years
CR3	5	5.224
SL1	0	3.871
SL2	0	2.064
TP3	2	5.763
TP4	1	5.586
Totals	8	22.508

Sometimes, data subsets can be split or combined in reasonable ways. For example, if the subsets were time periods, the data could be partitioned into decades, years, or months. The finer the division of the cells, the more sparse the data become within the cells. Too fine a partition allows random variation to dominate within each cell, but too coarse a partition may hide variation that is present within individual cells. In the present simple example, the most reasonable partition is into plants. Analysis of more complicated data sets may require examination of many partitionings.

First, a graphical technique is given to help the analyst understand what the data set shows. Then, a formal statistical procedure is presented to help quantify the strength of the evidence for patterns seen in the graphical investigation.

6.2.3.1.1 Graphical Technique

To explore the relations between cells, identify the cells on one axis. Then, for each cell, plot a point estimate of θ and an interval estimate of θ against the other axis. Patterns such as trends, outliers, or large scatter are then made visible.

In Example 6.4, the cells are plants. The data set from each plant was analyzed separately, using the tools of Section 6.2.1. The graph in Figure 6.21 shows the maximum likelihood estimate and a confidence interval for each plant, plotted side by side. For this handbook, the plot was produced with a graphics software package, although a hand-drawn sketch would be adequate to show the results.

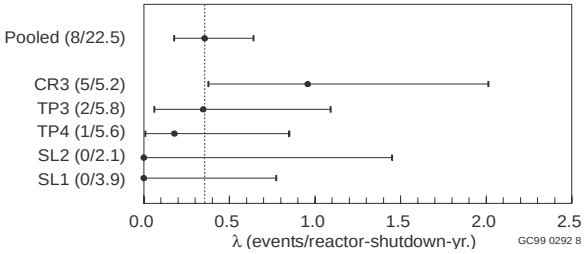


Figure 6.21 MLEs and 90% confidence intervals for θ, based on each plant's data and based on pooled data from all the plants.

The confidence interval for the pooled data is also shown. Take care, however: this interval is only valid if all the plants have the same θ, which is what must be decided. Nevertheless, the interval and point estimate for the pooled data give a useful reference for comparisons with the individual plants. For this reason, a vertical dotted line is drawn through the mean of the pooled data.

Note that the plants are not displayed in alphabetical order, which is a meaningless order for the event rate, but in order of decreasing $\hat{\lambda}$. (When two plants have the same MLE, as do SL1 and SL2, the upper confidence limit is used to determine the order.) Experience has shown that such a descending order assists the eye in making comparisons.

CR3 appears somewhat high compared to the others. Although there is considerable overlap of the intervals, the lower confidence limit for CR3 is just

barely higher than the MLE for the utility as a whole. Of course, the picture might give a different impression if slightly different intervals were used: 95% confidence intervals instead of 90% confidence intervals, or Bayes intervals with the Jeffreys noninformative prior instead of confidence intervals. From the graph alone, it is difficult to say whether the data can be pooled.

A graph like this should not be used to draw conclusions without also using a formal statistical test. For example, if many confidence intervals are plotted, based on data sets generated by the same θ, a few will be far from the others because of randomness alone. This was seen in Figure 6.4, where all the variation was due to randomness of the data, and some intervals did not overlap some others at all. Thus, an outlying interval does not prove that the θs are unequal. This same statement is true if other intervals are used, such as Bayes credible intervals based on the noninformative prior. The issue is the random variability of data, not the kind of interval constructed.

Conversely, if there are only a few intervals, intervals that just barely overlap can give strong evidence for a difference in the θs.

To quantify the strength of the evidence against poolability, a formal statistical procedure is given in the next subsection. The graph gives an indication of what the test might show, and helps in the interpretation of the test results. If the statistical test turns out to find a statistically significant difference between plants, it is natural then to ask what kind of difference is present. Figure 6.21 shows that most of the plants appear similar, with only one possible outlier. An unusually long interval, such as that seen in Figure 6.21 for SL2, is generally associated with a smaller exposure time. The picture provides insight even though it does not give a quantitative statistical test.

6.2.3.1.2 Statistical Test

The Chi-Squared Test. To study whether the rate is the same for different cells, use a **chi-squared test**. Many statistics texts, such as Bain and Engelhardt (1992, Chapter 13), discuss this test, and many software packages perform the chi-squared test. It is presented here in enough detail so that the reader could perform the calculations by hand if necessary, because it is instructive to see how the test works.

Let the null hypothesis be:

H_0: θ is the same in all the data subsets.

In the present application, the data subsets are the five plants. The method is to see what kind of data would be expected when θ really is constant, and then to see how much the observed counts differ from the expected counts. If the difference is small, the counts are consistent with the hypothesis H_0 that the rate is constant. If, instead, the difference is large, the counts show strong evidence against H_0.

Write x_j and t_j for the count and exposure time corresponding to the jth cell, and let $x = \Sigma x_j$ and $t = \Sigma t_j$. If H_0 is true, that is, if θ is really the same for all the plants, then the estimate (MLE) of θ is $\hat{\lambda} = x/t$. The estimate of the expected count is built from this quantity. Assuming the hypothesis of a single rate θ, an estimate of the expected count for the jth cell is simply:

$$e_j = \hat{\lambda} t_j .$$

In Example 6.4, the estimate of θ is 8/22.508 = 0.355 events per shutdown-year. Therefore, the expected count for CR3 is the estimate of θ times the exposure time for CR3, 0.335 × 5.224 = 1.857 events. Table 6.6 is an extension of the original table given in Example 6.4, showing the quantities needed for the calculation.

Table 6.6 Quantities for calculation of chi-squared test.

Cell code	x_i	t_i	e_i
CR3	5	5.224	1.857
SL1	0	3.871	1.376
SL2	0	2.064	0.734
TP3	2	5.763	2.048
TP4	1	5.586	1.985
Totals	8	22.508	8.000

The total of the expected counts agrees with the total of the observed counts, except possibly for small round-off error.

The test for equality of rates that is considered here is based on the following calculated quantity,

$$X^2 = \Sigma_j (x_j - e_j)^2 / e_j ,$$

sometimes called the **Pearson chi-squared statistic**, after its inventor, Karl Pearson, or simply the **chi-squared statistic**. The notation became standard long before the custom developed of using upper-case letters for random variables and lower-case letters for numbers. In the discussion below, the context must reveal whether X^2 refers to the random variable or the observed value.

Observe that X^2 is large if the x_js (observed counts) differ greatly from the e_js (expected values when H_0 is true). Conversely, X^2 is small if the observed values are close to the expected values. This statement is made more precise as follows. When H_0 is true and the total count is large, the distribution of X^2 has a distribution that is approximately chi-squared with c ! 1 degrees of freedom, where c is the number of cells. If the calculated value of X^2 is large compared to the chi-squared distribution, there is strong evidence that H_0 is false; the larger the X^2 value, the stronger the evidence.

For the data of Table 6.4, X^2 = 7.92, which is the 90.6th percentile of the chi-squared distribution with four degrees of freedom. The next subsection discusses the interpretation of this.

Interpretation of Test Results. Suppose, for any example with 5 cells, that X^2 were 9.8. A table of the chi-squared distribution shows that 9.488 is the 95th percentile of the chi-squared distribution with 4 degrees of freedom, and 11.14 is the 97.5th percentile. After comparing X^2 to these values, we would conclude that the evidence is strong against H_0, but not overwhelming. The full statement is:

- If H_0 is true, that is, if all the cells have the same θ, the chance of seeing such a large X^2 is less than 0.05 but more than 0.025.

Common abbreviated ways of saying this are:

- We reject H_0 at the 5% **significance level**, but not at the 2.5% significance level.
- The difference between cells is **statistically significant at the 0.05 level**, but not at the 0.025 level.
- The **p-value** is between 0.05 and 0.025.

There will be some false alarms. Even if θ is exactly the same for all the cells, sometimes X^2 will be large, just from randomness. It will be greater than the 95th percentile for 5% of the data sets, and it will be greater than the 99th percentile for 1% of the data sets. If we observed such a value for X^2, we would probably decide that the data could not be pooled. In that case, we would have believed a false alarm and made the incorrect decision. Just as with confidence intervals, we cannot be sure that this data set is not one of the rare unlucky ones. But following the averages leads us to the correct decision most of the time.

If, instead, X^2 were 4.1, it would be near the 60th percentile of the chi-squared distribution, and therefore be in the range of values that would be expected under H_0. We would say the observed counts are consistent with the hypothesis H_0, or H_0 cannot be rejected, or the evidence against H_0 is weak. We would not conclude that H_0 is true, because it probably is not exactly true to the tenth decimal place, but the conclusion would be that H_0 cannot be rejected by the data.

In fact, for the data of Table 6.6, X^2 equals 7.92, which is the 90.6th percentile of the chi-squared distribution with 4 degrees of freedom. That means: if all five plants have the same event rate, there is a 9.4% probability of seeing such a large value of X^2. The evidence against H_0 is not convincingly strong. CR3 might be suspected of having a higher event rate, but the evidence is not strong enough to prove this.

The traditional cut-off is 5%. The difference between cells is called **statistically significant**, with no qualifying phrase, if it is significant at the 0.05 level. This is tradition only, but it is very widely followed.

In actual data analysis, do not stop with the decision that a difference is, or is not, statistically significant. Do not even stop after reporting the p-value. That may be acceptable if the p-value is very small (much less than 0.05) or very large (much larger than 0.05). In many cases, however, statistical significance is far from the whole story. Engineering significance is just as important.

To illustrate this, consider a possible follow-up to the above statistical analysis of Example 6.4. As mentioned, the statistical evidence against poolability is not strong, but some might consider it borderline. Therefore, a thorough analysis would ask questions such as:

- Are there engineering reasons for expecting CR3 to have a different event rate than the other plants do, either because of the hardware or because of procedures during shutdown? (Be warned that it is easy to find justifications in hindsight, after seeing the data. It might be wise to hide the data and ask these questions of a different knowledgeable person.)
- What are the consequences for the PRA analysis if the data are pooled or if, instead, CR3 is treated separately from the other plants? Does the decision to pool or not make any practical difference?

Required Sample Size. The above considerations are valid if the total count is "large," or more precisely, if the e_js are "large." If the e_js are small, the chi-squared distribution is not a good approximation to the distribution of X^2. Thus, the user must ask how large a count is necessary for the chi-squared approximation to be adequate. An overly conservative rule is that each expected cell-count, e_j, should be 5.0 or larger. Despite its conservatism, this rule is still widely used, and cited in the statistical literature and by some software packages.

A readable discussion of chi-squared tests by Moore (1986, p.71) is applicable here. Citing the work of Roscoe and Byars (1971), the following recommendations are made:

(1) With equiprobable cells, the average expected frequency should be at least 1 when testing at the 0.05 level. In other words, use the chi-squared approximation at the 5% level when $x/c \geq 1$, where x is the number of events and c is the number of cells. At the 1% level, the chi-squared approximation is recommended if $x/c \geq 2$.

(2) When the cells are not approximately equiprobable, the average expected frequencies in (1) should be doubled. Thus, the recommendation is that at the 5% level $x/c \geq 2$, and at the 1% level $x/c \geq 4$.

Note that in rules (1) and (2) above, the recommendation is based on the average rather than the minimum expected cell-count. As noted by Koehler and Larntz (1980), any rule such as (2) may be defeated by a sufficiently skewed assignment of cell probabilities.

Roscoe and Byars also recommend when $c = 2$ that the chi-squared test should be replaced by the test based on the exact binomial distribution of X_1 conditional on the total event count. For example, if the two cells had the same exposure times, we would expect that half of the events would be generated in each cell. More generally, if

- the two cells have exposure times t_1 and t_2,
- a total of x events are observed, and
- θ is the same for both cells,

then, conditional on x, X_1 has a binomial(n, p) distribution, with $p = t_1/(t_1 + t_2)$. Exact binomial tests are discussed by Bain and Engelhardt (1992, p.405).

Example 6.4 has $x = 8$ and $c = 5$. The cells are not equiprobable, that is, e_j is not the same for all cells, because the plants did not all have the same exposure time. Nevertheless, the expected cell counts differ from each other by, at most, a factor of two. This is not a large departure from equiprobability, as differences of an order of magnitude would be. Because $x/c = 1.6$, and the calculated significance level is about 10%, the sample size is large enough for the chi-squared approximation to be adequate. The conclusions reached earlier still stand. If, on the other hand, the sample size had been considerably smaller, one would have to say that the p-value is *approximately* given by the chi-squared distribution, but that the exact p-value has not been found.

If the expected cell-counts are so small that the chi-squared approximation is not recommended, the analyst can pool data in some "adjacent cells," thereby increasing the expected cell-counts.

In the Example 6.4, suppose that there were engineering reasons for thinking that the event rate is similar at units at a single site. Then, the sister units might be pooled, transforming the original table of Example 6.4 into Table 6.7 here.

Table 6.7 Shutdown LOSP events at three sites, 1980-96.

Site code	Events	Plant shutdown years
CR	5	5.224
SL	0	5.935
TP	3	11.349

We repeat, this pooling of cells is not required with the actual data, but it could be useful if (a) the cell counts were smaller and (b) there were engineering reasons for believing that the pooled cells are relatively homogeneous, that is, the event rates are similar for both units at a site, more similar than the event rates at different sites.

Generally speaking, a chi-squared test based on a larger number of cells will have better power for detecting when rates are not equal, but this also makes it more difficult to satisfy guidelines on expected cell-counts for the chi-squared approximation. Thus, it is sometimes necessary to make a compromise between expected cell counts and the number of cells.

Options involving the exact distribution of X^2 are also possible. The most widely known commercial software for calculating the exact p-value is StatXact (1999).

6.2.3.2 No Time Trend

The chi-squared method given above does not use any ordering of the cells. Even if the test were for differences in years, say, the test would not use the natural ordering by calendar year or by plant age. When there is a meaningful order to the data subsets, it may be useful to perform additional analyses. The analysis given above is valid, but an additional possible analysis, making use of time order, is considered now.

The methods will be illustrated with Example 6.5.

6.2.3.2.1 Graphical Techniques

Confidence-Interval Plot. First, the same kind of plot that was used in the previous subsection can be used here. The time axis is divided into cells, or **bins** in the terminology of some authors. For example, if the time span is divided into calendar years, the counts and reactor-critical-years for Example 6.5 are given in Table 6.8.

Example 6.5 Unplanned HPCI demands.

Grant et al. (1995, Table B-5) list 63 unplanned demands for the HPCI system to start at 23 BWRs during 1987-1993. The demand dates are given in columns below, in format MM/DD/YY.

01/05/87	08/03/87	03/05/89	08/16/90	08/25/91
01/07/87	08/16/87	03/25/89	08/19/90	09/11/91
01/26/87	08/29/87	08/26/89	09/02/90	12/17/91
02/18/87	01/10/88	09/03/89	09/27/90	02/02/92
02/24/87	04/30/88	11/05/89	10/12/90	06/25/92
03/11/87	05/27/88	11/25/89	10/17/90	08/27/92
04/03/87	08/05/88	12/20/89	11/26/90	09/30/92
04/16/87	08/25/88	01/12/90	01/18/91	10/15/92
04/22/87	08/26/88	01/28/90	01/25/91	11/18/92
07/23/87	09/04/88	03/19/90	02/27/91	04/20/93
07/26/87	11/01/88	03/19/90	04/23/91	07/30/93
07/30/87	11/16/88	06/20/90	07/18/91	
08/03/87	12/17/88	07/27/90	07/31/91	

Table 6.8 HPCI demands and reactor-critical-years.

Calendar year	HPCI demands	Reactor-critical-years
1987	16	14.63
1988	10	14.15
1989	7	15.75
1990	13	17.77
1991	9	17.11
1992	6	17.19
1993	2	17.34

This table has the same form as in Example 6.4, showing cells with events and exposure times. The relevant exposure time is reactor-critical-years, because the HPCI system uses a turbine-driven pump, which can only be demanded when the reactor is producing steam. The counts come from the tabulated events of Example 6.5, and the critical-years can be constructed from information in Poloski et al. (1999a). The variation in critical-years results from the facts that several reactors were shut down for extended periods, and one reactor did not receive its low power license until 1989.

This leads to a plot similar to Figure 6.21, showing the estimated value of the demand frequency, δ, and a confidence interval for each year. This is shown in Figure 6.22.

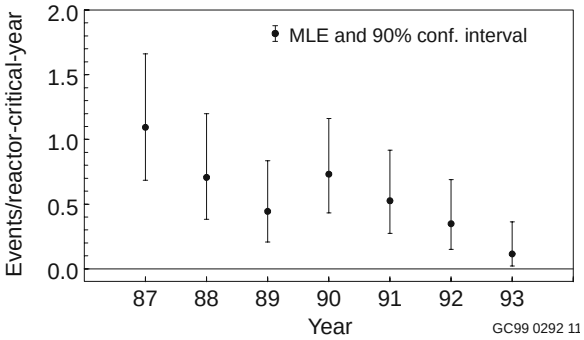


Figure 6.22 MLEs and 90% confidence intervals for δ, each based on data from one calendar year.

Figure 6.22 seems to indicate a decreasing trend in the frequency of HPCI demands. However, the picture does not reveal whether the apparent trend is perhaps merely the result of random scatter. To answer that question, a formal statistical test is necessary, quantifying the strength of the evidence. Such tests will be given in Section 6.2.3.2.2.

Cumulative Plot. Figure 6.22 required a choice of how to divide the time axis into cells. A different plot, given next, does not require any such choice, if the dates of the events are recorded. Plot the cumulative event count at the n event dates.

Figure 6.23 shows this for Example 6.5. The events are arranged in chronological order, and the cumulative count of events is plotted against the event times.

The **slope** of a string of plotted points is defined as the vertical change in the string divided by the horizontal change, $\Delta y / \Delta x$. This is the familiar definition of slope from mathematics courses. In the plot given here, the horizontal distance between two points is elapsed time,

and the vertical distance is the total number of events that occurred during that time period. Therefore,

slope = (number of events)/(elapsed time) ,

so the slope is a graphical estimator of the event frequency, θ. A constant slope, or a straight line, indicates a constant θ. Changes in slope indicate changes in θ: if the slope becomes steeper, θ is increasing, and if the slope becomes less steep, θ is decreasing.

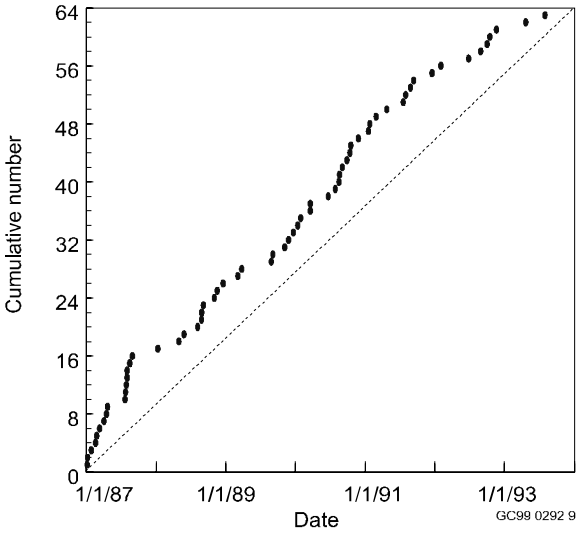


Figure 6.23 Cumulative number of HPCI demands, by date.

In Example 6.5 the time axis represents calendar years. Because the relevant frequency is events per reactor-critical-year, it would be better to plot the time axis in terms of total reactor-critical-years from the start of 1987. However, it is somewhat difficult to calculate the reactor-critical-years preceding any particular event, or equivalently, the reactor-critical-years between successive events. Therefore, simple calendar years are used. This is adequate if the number of reactors operating at any time is fairly constant, because then the rate per reactor-critical-year remains roughly proportional to the rate per industry-calendar year. In the present case, as shown by Table 6.8, later calendar-years correspond to more critical-years than do early calendar- years.

The slope in Figure 6.23 is steepest on the left, and gradually lessens, so that the plot is rising fastest on the left and more gently on the right. More HPCI demands are packed into a time interval on the left than into a time interval of the same length on the right. This indicates that the frequency of unplanned HPCI demands was decreasing during the time period of the study. Thus, this figure leads to the same general conclusion as does Figure 6.22. Figure 6.23 shows more detail, with the individual events plotted, but it is less accurate in this example because we have not gone through the work of plotting events versus reactor-critical time.

It is important that the horizontal axis cover the entire data-collection period and not stop at the final event. In Figure 6.23, the lack of events during the last half of 1993 contributes to the overall curvature of the plot.

If the frequency is constant, the plot should follow a roughly straight line. For comparison, it is useful to show a straight diagonal line, going from height 0 at the start of the data collection period to height $n + 1$ at the end of the data collection period, where n is the number of data points.

In Figure 6.23, the diagonal line is shown as a dotted line, rising from height 0 on the left to height $n + 1 = 64$ on the right.

As mentioned above, the early calendar years contain fewer reactor-critical-years than do the later calendar years. Therefore, the time axis in Figure 6.23 would reflect reactor-critical-years more accurately if the left end of the axis were compressed slightly or the right end were stretched slightly. The effect would be to increase the curvature of the plot, making it rise more quickly on the left and more slowly on the right.

A cumulative plot contains random bounces and clusters, so it is not clear whether the observed pattern is more than the result of randomness. As always, a formal statistical test will be needed to measure the strength of the evidence against the hypothesis of constant event frequency.

6.2.3.2.2 Statistical Tests for a Trend in θ

The Chi-Squared Test. This is the same test as given in Section 6.2.3.1.2, only now the cells are years or similar divisions of time.

In Example 6.5, the p-value is 0.009, meaning that a random data set with constant θ would show this much variability with probability only 0.9%. Two points are worth noting.

- The chi-squared test makes no use of the order of the cells. It would give exactly the same conclusion if the intervals in Figure 6.22 were scrambled in a random order instead of generally decreasing from left to right.

- The calculated p-value is accurate enough to use, by the guidelines of Section 6.2.3.1.2, because the number of events is 63, and the number of cells is 7, so x/c = 63/7 = 9. Even splitting the cells into six-month periods or smaller periods would be justified.

Chapter 7 will take Figure 6.22, fit a trend, and perform an additional test based on the fit; see Sections 7.2.3 and 7.2.4. Therefore, the chi-squared test is not discussed further here.

The Laplace Test. This test does not use the binning of times into cells, but instead uses the exact dates. In the example, there are 63 occurrences of events during a seven-year period. In general, consider a time interval [0, L], and suppose that during this period n events occur at successive random times $T_1, T_2, \ldots, T_n$. Although the number of occurrences, n, is random when the plants are observed for a fixed length of time L, we condition on the value of n, and so treat it as fixed. Consider the null hypothesis:

H_0: θ is constant over time.

Consider the alternative hypothesis:

H_1: θ is either an increasing or a decreasing function of time.

This hypothesis says that the events tend to occur more at one end of the interval than at the other. A test that is often used is based on the mean of the failure times, $\overline{T} = \Sigma_i T_i / n$. The intuitive basis for the test is the following. If θ is constant, about half of the events should occur before time $L/2$ and half afterwards, and the average event time should be close to $L/2$. On the other hand, if θ is decreasing, more events are expected early and fewer later, so the average event time should be smaller than $L/2$. Similarly, if θ is increasing, the average event time is expected to be larger than $L/2$. Therefore, the test rejects H_0 if $\overline{T}$ is far from $L/2$. Positive values of the difference $\overline{T} - L/2$ indicate an increasing trend, and negative values indicate a decreasing trend.

When H_0 is true, $\overline{T}$ has expected value $L/2$ and variance $L^2/(12n)$. The resulting test statistic is

$$U = \frac{\overline{T} - L/2}{L/\sqrt{12n}} .$$

The statistic U is approximately standard normal for $n \geq 3$. A test of H_0 at significance level 0.05 versus an increasing alternative,

H_1: θ is increasing in time,

would reject H_0 if $U \geq 1.645$. A 0.05 level test versus a decreasing alternative,

H_1: θ is decreasing in time,

would reject H_0 if $U \leq -1.645$. Of course, ±1.645 are the 95th and 5th percentiles, respectively, of the standard normal distribution. A two-sided test, that is, a test against the original two-sided alternative hypothesis, at the 0.10 level would reject H_0 if $|U| \geq 1.645$.

This test, generally known as the "Laplace" test, is discussed by Cox and Lewis (1978, p. 47). The Laplace test is known to be good for detecting a wide variety of monotonic trends, and consequently it is recommended as a general tool for testing against such alternatives.

Let us apply the Laplace test to the HPCI-demand data of Example 6.5. First, the dates must be converted to times. The first event time is 0.011 years after January 1, 1987, the final event is 6.581 years after the starting date, and the other times are calculated similarly. Here, a "year" is interpreted as a 365-day year. The total number of 365-day years is L = 7.00. The mean of the event times can be calculated to be 2.73. Therefore, the calculated value of U is

$$\frac{2.73 - 3.5}{7.00/\sqrt{12 \times 63}} = -3.02 .$$

This is statistically very significant. The value 3.02 is the 0.1th percentile of the standard normal distribution. Thus, the evidence is very strong against a constant demand rate, in favor instead of a decreasing demand rate. Even against the two-sided hypothesis

H_1: θ is increasing or decreasing in time,

the p-value is Pr($|U| > 3.02$) = 0.002.

In the example, the Laplace test statistic was calculated in terms of calendar time instead of reactor-critical-time. As remarked earlier, using reactor-critical-time would increase the curvature of the plot in Figure 6.23. A similar argument shows that using reactor-critical-time in computing U would increase

the strength of the evidence against the hypothesis of a constant demand rate. However, the computations would be very tedious. That is an advantage of the chi-squared test, because it is typically easier to find the exact relevant exposure time for blocks of time, such as years, than for each individual event.

In the example, the result of the Laplace test agrees with the result from the chi-squared test, but is more conclusive. The chi-squared test gave a p-value of 0.009, meaning that if H_0 is true, the cells would appear so different from each other with probability only 0.009. The Laplace test gives a p-value of 0.002.

The chi-squared and Laplace tests differ because they are concerned with different alternatives to H_0. The chi-squared test is concerned with any variation from cell to cell (from year to year in the example). If the event rate goes up and down erratically, that is just as much evidence against H_0 as if the event rate decreases monotonically. The Laplace test, on the other hand, is focused on the alternative of a trend. It has more power for detecting trends, but no power at all for detecting erratic changes upward and downward.

Other tests exist in this setting. See Ascher and Feingold (1984, page 80) and Engelhardt (1994, p. 19) for details.

6.2.3.3 No Multiple Failures

The second assumption of the Poisson process is that there are no exactly simultaneous failures. In practice this means that common-cause failures do not occur. In most situations, common-cause failures will occur from time to time. This was seen in some of the examples discussed in Section 2.2. However, if common-cause events are relatively infrequent, their effect on the validity of the Poisson model can normally be ignored.

No statistical methods are given here to examine whether common-cause events can occur. Instead, the analyst should think of the engineering reasons why common-cause events might be rare or frequent, and the data should be examined to discover how frequent common-cause events are in practice.

In Example 6.5, HPCI demands, it is reasonable that common-cause events could occur only at multiple units at a single site. There was one such pair of events in the data, with HPCI demands at Hatch 1 and Hatch 2, both on 08/03/87. Examination of the LERs reveals that the demands occurred from different causes. They happened at different times, and so were not exactly simultaneous. The conclusion is that common causes may induce exactly simultaneous events, but they are infrequent.

If common-cause events are relatively frequent, so that they cannot be ignored, it might be necessary to perform two analyses, one of the "independent", or not-common-cause, events, and one of the common-cause occurrences. The frequency of independent events could be estimated using the methods given here. The common cause events would have to be analyzed by other methods, such as methods described in the references given in Section 1.3.

6.2.3.4 Independence of Disjoint Time Periods

This section is less important than the others, and of interest only to truly dedicated readers. Others should skip directly to Section 6.2.3.5.

The final assumption of the Poisson model is that event occurrences in disjoint time periods are statistically independent. This should first be addressed by careful thinking, similar to that in the examples of Section 2.2. However, the following statistical approach may also be useful.

One possible type of dependence would be if events tend to cluster in time: large between-event times tend to occur in succession, or similarly small ones tend to occur in succession. For example, suppose that a repair is done incorrectly several times in succession, leading to small times between failures. The occurrence of a failure on one day would increase the probability of a failure in the next short time period, violating the Poisson assumption. After the problem is diagnosed, the personnel receive training in proper repair procedures, thereafter resulting in larger times between failures.

To illustrate the ideas, an example with no trend is needed. The shutdown LOSP events introduced in Section 2.2 can be used as such an example. The data are restricted here to the years 1991-1996, primarily to reduce any effect of the overall downward trend in total shutdown tome. Atwood et al. (1998) report 24 plant-centered LOSP events during shutdown in 1991-1996. They are given as Example 6.6.

The null hypothesis is that the successive times between events are independent and exponentially distributed. We consider the alternative hypotheses that

- the times are not exponentially distributed, possibly with more short times between events than expected from an exponential distribution; or

- successive times are correlated, that is that short times tend to be followed by short times and long times by long times.

Example 6.6 Dates of shutdown LOSP events and days between them.

The consecutive dates of shutdown LOSP events are shown in columns below. After each date is the time since the preceding event, in days. For the first event, the time since the start of the study period is shown. Also, the time is shown from the last event to the end of the study period, a 25th "between-event time."

03/07/91	66	04/02/92	10	09/27/94	129
03/13/91	6	04/06/92	4	11/18/94	52
03/20/91	7	04/28/92	22	02/27/95	101
04/02/91	13	04/08/93	345	10/21/95	236
06/22/91	81	05/19/93	41	01/20/96	91
07/24/91	32	06/22/93	34	05/23/96	124
10/20/91	88	06/26/93	4	—	223
01/29/92	101	10/12/93	108		
03/23/92	54	05/21/94	221		

Section 6.6.2.3 discusses ways to investigate whether data come from a particular distribution. Therefore, the issue of the exponential distribution is deferred to that section. The issue of serial correlation motivates the following procedure. Let y_i be the ith time between events, and let x_i be the $(i-1)$ time between events, $x_i = y_{i-1}$. We look to see if x_i and y_i are correlated.

In the above example, the first few (x, y) pairs are (66, 6), (6, 7), and (7, 13), and the final pair is (124, 223).

6.2.3.4.1 Graphical Method

As just mentioned, the issue of whether the distribution is exponential is deferred to Section 6.6.2.3. Consider here the question of serial correlation. A scatter plot of x versus y will indicate whether the values are correlated. However, with skewed data the large values tend to be visually dominant, distorting the overall message of the plot. One could try an ad hoc transformation, such as the logarithmic transformation, but a more universally applicable approach is to use the **ranks** of the variables. That is, sort the n times in increasing order, and assign rank 1 to the smallest time and rank n to the largest time.

In the example, the two shortest times are each equal to 4 days. Each is assigned the average of ranks 1 and 2, namely 1.5. The next largest time is 6 days, which is assigned rank 3, and so forth. The 17th and 18th times are each 101 days, so those two are each assigned rank 17.5. Selected values of x, y and their ranks are shown in Table 6.9. For compactness, not all of the values are printed.

Table 6.9 Calculations for analyzing LOSP dates.

x	rank(*x*)	*y*	rank(*y*)
—	—	66	13
66	13	6	3
6	3	7	4
7	4	13	6
13	6	81	14
81	14	32	8
32	8	88	15
88	15	101	17.5
101	17.5	54	12
54	12	10	5
...	...	...	...
52	11	101	17.5
101	17.5	236	24
236	24	91	16
91	16	124	20
124	20	223	23
223	23	—	—

Figure 6.24 shows a scatter plot of rank(x) versus rank(y). The plot seems to show very little pattern, indicating little or no correlation from one time to the next. The barely perceptible trend from lower left to upper right ("southwest to northeast") is probably not meaningful, but a hypothesis test will need to be performed to confirm or refute that judgment.

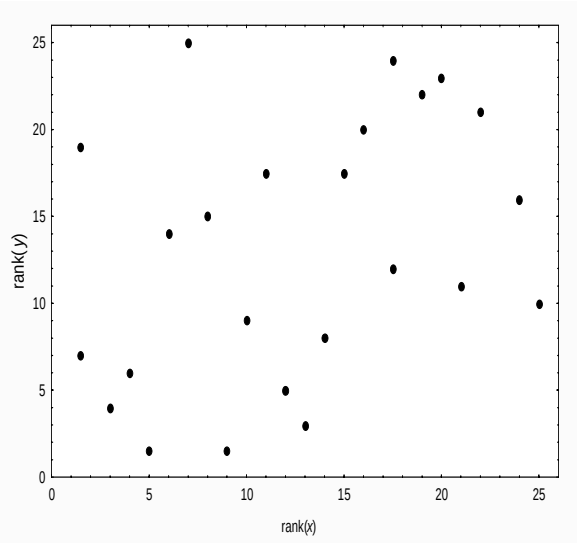


Figure 6.24 Scatter plot of rank(x) versus rank(y).

6.2.3.4.2 Statistical Tests

This section considers whether the between-event times are serially correlated. The question of whether they are exponentially distributed is discussed in Section 6.6.2.3, under the topic of goodness-of-fit tests.

To test for correlation, it is not appropriate to assume normality of the data. Instead, a nonparametric test should be used, that is, a test that does not assume any particular distributional form. A test statistic that is commonly produced by statistical software is Kendall's tau (J). Tau is defined in Conover (1999), Hollander and Wolfe (1999), and other books on nonparametric statistics.

Based on the data of Table 6.9, the hypothesis of no correlation between X and Y was tested. Kendall's tau gave a p-value of 0.08. This calculation indicates that the very slight trend seen in Figure 6.24 is not statistically significant.

Recall, from the discussion of Section 6.2.3.1.2, that a small p-value is not the end of an analysis. The p-value for this example, although larger than the customary cut-off of 0.05, is fairly small. This indicates that the trend in Figure 6.24 is somewhat unlikely under the assumption of no correlation. If we are concerned about this fact, we must seek possible engineering mechanisms for the trend. The data are times between LOSP events in the industry as a whole. Therefore, the most plausible explanation is the overall industry trend of fewer shutdown LOSP events. This trend would produce a tendency for the short times to occur together (primarily near the start of the data collection period), and the long times to occur together (primarily near the end of the data period).

6.2.3.5 Consistency of Data and Prior

As an example, if the prior distribution has mean $E_{\text{prior}}(\delta)$, but the observed data show x/t very different from the prior mean, the analyst might wonder if the data and the prior are consistent, or if, instead, the prior distribution was misinformed. To investigate this, one could ask what the prior probability is of getting the observed data. Actually, any individual x may have small probability, so a slightly more complicated question is appropriate.

Suppose first that x/t is in the right tail of the prior distribution. The relevant quantity is the prior probability of observing x or more events. This is

$$\Pr(X \geq x) = \int \Pr(X \geq x|\lambda)\, f_{\text{prior}}(\lambda) d\lambda \qquad (6.6)$$

where

$$\Pr(X \geq x|\lambda) = \sum_{k=x}^{\infty} e^{-\lambda t}(\lambda t)^k / k! = 1 - \sum_{k=0}^{x-1} e^{-\lambda t}(\lambda t)^k / k! \qquad (6.7)$$

In general, Equation 6.6 does not have a direct analytical expression. However, in the special case when the prior distribution is gamma($\alpha_{\text{prior}}, \beta_{\text{prior}}$), it can be shown that the probability in question equals

$$\Pr(X \geq x) = 1 - \sum_{k=0}^{x-1} \frac{\Gamma(\alpha + k)}{k!\,\Gamma(\alpha)} (t/\beta)^k (1 + t/\beta)^{-(\alpha+k)} \qquad (6.8)$$

where $\Gamma(s)$ is the gamma function, a generalization of the factorial function as described in Appendix A.7.6. The distribution defined by Equation 6.8 is named the **gamma-Poisson** or **negative binomial distribution**. The above probability can be evaluated with the aid of software.

When Equation 6.8 is not applicable, one method of approximating the integral in Equation 6.6 is by Monte Carlo sampling. Generate a large number of values of δ from the prior distribution. For each value of δ, let y be the value of Equation 6.7, which can be calculated directly. The average of the y values is an approximation of the integral in Equation 6.6. Another method of approximating the Equation 6.6 is by numerical integration.

If the probability given by Equation 6.6 is small, the observed data are not consistent with the prior belief — the prior belief mistakenly expected δ to be smaller than it apparently is. When should the probability be considered "small"? Many people consider probabilities < 0.05 to be "small," but there is no rigid rule.

Similarly, if x/t is in the left tail of the prior distribution, the relevant quantity is the prior probability that $X \# x$. When the prior is a gamma distribution, the desired probability is the analogue of the sum in Equation 6.8,

$$\Pr(X \leq x) = \sum_{k-0}^{x} \frac{\Gamma(\alpha+k)}{k!\,\Gamma(\alpha)} (t/\beta)^k (1+t/\beta)^{-(\alpha+k)}$$

In any case, the desired probability can be approximated by Monte Carlo sampling. If that probability is small, the prior distribution mistakenly expected δ to be larger than it apparently is.

In Example 6.3, we ask whether the observed zero failures in 2102 reactor-calendar-years is consistent with the WASH-1400 prior, lognormal with median 1E-3 per year and error factor 10. To investigate this, 100,000 random values of θ were generated from the lognormal prior. (The details are given below.) For each θ, $\Pr(X \# 0) = \exp(!2102\theta)$ was found. The mean of these probabilities was 0.245. This is a sample mean, and it estimates the true probability. It is not small, and therefore gives no reason to question the applicability of the prior.

One must ask whether the sample was large enough. The software that calculated the sample mean also calculated the standard error to be 0.0009. Recall from Section 6.2.1.2 that in general a 95% confidence interval can be approximated as the estimate plus or minus 2×(standard error). In this case, this interval becomes 0.245 ± 0.002. We conclude that the true mean equals 0.245 except perhaps for random error in the third digit. This shows that the sample size was more than large enough to give an answer to the accuracy required.

The recipe for generating θ from a lognormal distribution is as follows:

(1) Generate z from a standard normal distribution, using commercial software,
(2) Define loglam = $\lambda + \sigma z$, where λ and σ were found in Section 6.2.2.7.2, and then
(3) Define lambda = exp(loglam).

6.3 Failures to Change State: Failure on Demand

This section is similar to Section 6.2, but the details are different. The structure of this section parallels that of Section 6.2 almost exactly, and some admonitions from that section are repeated here. The most important topics for a first-time reader are:

- Maximum likelihood estimation (6.3.1.1),
- Bayesian estimation, especially with a discrete prior or a conjugate prior (6.3.1-6.3.2.3), and
- Model validation, especially using graphical tools (portions of 6.3.3).

This section applies to data satisfying the assumptions of Section 2.3.2.1. The probability of a failure on demand is denoted p, a unitless quantity. The data consist of x failures in n demands, with $0 \# x \# n$. Before the data are generated, the number of failures is random, denoted X. For any particular number x, the probability of x failures in n demands is

$$\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x} , \qquad (6.9)$$

where the binomial coefficient is defined as

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} .$$

The methods will be illustrated by the following hypothetical data set.

Example 6.7 AFW turbine-train failure to start

> In the last 8 demands of the turbine train of the auxiliary feedwater (AFW) system at a PWR, the train failed to start 1 time. Let p denote the probability of failure to start for this train.

As in Section 6.2, frequentist methods are presented first, followed by Bayesian methods. This choice is made because the frequentist point estimate is so very simple, not because frequentist estimation is preferable to Bayesian estimation. Indeed, in PRA p is normally estimated in a Bayesian way.

6.3.1 Frequentist or Classical Estimation

6.3.1.1 Point Estimate

The most commonly used frequentist estimate is the **maximum likelihood estimate** (MLE). It is found by taking the **likelihood**, given by Equation 6.9, and treating it as a function of p. The value of p that maximizes the likelihood is called the MLE. It can be shown, by setting a derivative to zero, that the maximum likelihood estimate (MLE) of p is $\hat{p} = x / n$. This is intuitively appealing – the observed number of failures divided by the observed number of demands.

Figure 6.25 shows the likelihood as a function of p, for the data of Example 6.7. The figure shows that the likelihood is maximized at p = 1/8, as stated by the formula.

If several subsets of data, such as data corresponding to several plants, several types of demand, or several years, are assumed to have the same p, data from the various sources may be combined, or **pooled**, for an overall estimate. Denoting the number of failures and demands in data subset j by x_j and n_j, respectively, let $x = \Sigma_j x_j$ and $n = \Sigma_j n_j$. The MLE is x/n.

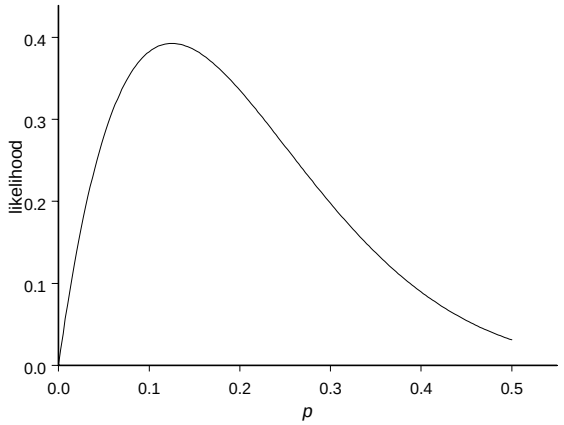


Figure 6.25 Likelihood as a function of p, for the data of Example 6.7.

As mentioned in Section 6.2.1.1, final answers will be shown in this handbook with few significant digits, to avoid giving the impression that the final answer reflects precise knowledge of the parameter. Intermediate values will show more significant digits, to prevent roundoff errors from accumulating.

6.3.1.2 Standard Deviation of Estimator

The number of failures is random. One number was observed, but if the demands were repeated a different number of failures might be observed. Therefore, the estimator is random, and the calculated estimate is the value it happened to take this time. Considering the data as random, one could write $\hat{P} = X / n$. This notation is consistent with the use of upper case letters for random variables, although it is customary in the literature to write $\hat{p}$ for both the random variable and the calculated value. The standard deviation of the estimator is $[p(1 - p)/n]^{1/2}$. Substitution of the estimate $\hat{p}$ for p yields an estimate of the standard deviation,

$$[\hat{p}(1-\hat{p})/n]^{1/2} .$$

The estimated standard deviation of an estimator is also called the **standard error** of the estimate. The handy rule given in Section 6.2.1.2 applies here as well:

MLE ± 2×(standard error)

is an approximate 95% confidence interval for p, when the number of demands, n, is large. However, an exact confidence interval is given below.

In Example 6.7, the standard error for p is

$[0.125 \times (1 - 0.125) / 8]^{1/2} = 0.12$.

6.3.1.3 Confidence Interval for *p*

Confidence intervals are used in two ways in this handbook. They give a standard of comparison, when Bayes credible intervals are found based on so-called noninformative priors, and they can be used (but are not required) in some plots for validating model assumptions. Therefore, readers may wish to skim the present section quickly on the first reading.

The interpretation of confidence intervals is given in Appendix B and in Section 6.2.1.3. It is so important that it is repeated here. In the frequentist approach, p is fixed and the data are random. Therefore the maximum likelihood estimator and the confidence limits are all random. For most data sets, the MLE, $\hat{p}$, will be close to the true value of p, and the confidence interval will contain p. Sometimes, however, the MLE will be rather far from p, and sometimes (less than 10% of the time) the 90% confidence interval will not contain p. The procedure is good in the sense that most of the time it gives good answers, but the analyst never knows if the current data set is one of the unlucky ones. A figure like Figure 6.4 could be constructed for p, to illustrate that many data sets could be generated from the same p, yielding many confidence intervals, *most of which* contain the true value of p.

The following material is drawn from Johnson et al. (1992, Section 3.8.3). A confidence interval for p can be expressed in terms of percentiles of a beta distribution. Appendix A.7.8 presents the basic facts about the beta distribution. As mentioned there, the beta family of distributions includes many distributions that are defined on the range from 0 to 1, including the uniform distribution, bell-shaped distributions, and U-shaped distributions. The beta distribution is also discussed more fully in the section below on Bayesian estimation.

Denote the lower and upper ends of a 100(1 ! γ)% confidence interval by $p_{\text{conf},\ \gamma/2}$ and $p_{\text{conf},\ 1!\ \gamma/2}$, respectively. It can be shown that the lower limit is

$$p_{\text{conf},\ \gamma/2} = \text{beta}_{\gamma/2}(x,\ n\ !\ x + 1)$$

and the upper limit is

$$p_{\text{conf},\ 1!\ \gamma/2} = \text{beta}_{1\ !\ \gamma/2}(x + 1,\ n\ !\ x)$$

where $\text{beta}_q(\alpha, \beta)$ denotes the q quantile, or $100\times q$ percentile, of the beta(α, β) distribution. For example, a 90% confidence interval for p is given by $\text{beta}_{0.05}(x, n - x + 1)$ and $\text{beta}_{0.95}(x + 1, n - x)$. If $x = 0$, the beta distribution for the lower limit is not defined; in that case, set $p_{\text{conf}, \alpha/2} = 0$. Similarly, if $x = n$, the beta distribution for the upper limit is not defined; in that case, set $p_{\text{conf}, 1 - \alpha/2} = 1$. In any case, note carefully that the parameters of the beta distribution are not quite the same for the lower and upper endpoints.

Appendix C tabulates selected percentiles of the beta distribution. However, interpolation may be required. Some software packages, including commonly used spreadsheets such as Microsoft Excel (2001) and Quattro Pro (2001), calculate the percentiles of the beta distribution. Those calculations are more accurate than interpolating tables. Finally, Appendix A.7.8 gives a last-resort method, which allows beta percentiles to be calculated by complicated formulas involving tabulated percentiles of the F distribution.

In Example 6.7, with 1 AFW train failure in 8 demands, suppose that a 90% interval is to be found. Then $\alpha = 0.10$, and $1 - \alpha/2 = 0.95$. For the lower limit, $\text{beta}_{0.05}(1, 8, 1+1) = 6.39\text{E}{-3}$, from Table C.5. Thus,

$p_{\text{conf}, 0.05} = 0.0064$.

For the upper limit, $\text{beta}_{0.95}(1+1, 8, 1) = 4.71\text{E}{-1}$, also from Table C.5. Thus,

$p_{\text{conf}, 0.95} = 0.47$.

6.3.2 Bayesian Estimation

Section 6.2.2.1 gives an overview of Bayesian estimation, which applies here. Just as for θ in that section, Bayesian estimation of p involves several steps. The prior belief about p is quantified by a probability distribution, the **prior distribution**. This distribution will be restricted to the range [0,1], because p must lie between 0 and 1, and it will assign the most probability to the values of p that are deemed most plausible. The data are then collected, and the **likelihood function** is constructed. The likelihood function is given by Equation 6.9 for failures on demand. It is the probability of the observed data, written as a function of p. Finally, the **posterior distribution** is constructed, by combining the prior distribution and the likelihood function through Bayes' theorem. The posterior distribution shows the updated belief about the values of p. It is a modification of the prior belief that accounts for the observed data.

Figure 6.5, showing the effect of various data sets on the posterior distribution, is worth studying. Although that figure refers to θ, exactly the same idea applies to p.

The subsections below consider estimation of p using various possible prior distributions. The simplest prior distribution is discrete. The posterior can be calculated easily, for example, by a spreadsheet. The next simplest prior is called **conjugate**; this prior combines neatly with the likelihood to give a posterior that can be evaluated by simple formulas. Finally, the most general priors are considered; the posterior distribution in such a case can only be found by numerical integration or by random sampling.

Section 6.2.2.2 discusses how to choose a prior, and gives references for further reading. It applies to estimation of p as much as to estimation of θ, and should be read in connection with the material given below.

6.3.2.1 Estimation with a Discrete Prior

The explanation here will be easier to follow if the examples in Section 6.2.2.3 have also been read. The parameter p is assumed to take one of only m possible values, $p_1, \ldots, p_m$. Denote the p.d.f. by f, so $f(p_i) = \Pr(p_i)$, the prior probability that the parameter has the value p_i. After evidence E is observed, Bayes' theorem says:

$$f(p_i|E) = \frac{f(p_i)L(E|p_i)}{\sum_{j=1}^{m} L(E|p_j)f(p_j)} \qquad (6.10)$$

where

$f(p_i \mid E)$ = the probability of p_i given evidence E (posterior distribution),

$f(p_i)$ = the probability of p_i prior to having evidence E (prior distribution), and

$L(E \mid p_i)$ = the likelihood function (probability of the evidence given p_i).

Just as in Section 6.2.2.3, the denominator in Equation 6.10, the total probability of the evidence E, is simply a normalizing constant.

When the evidence is in the form of x failures in n demands and the assumptions for a binomial distribu-

tion are satisfied, the likelihood function is the given by Equation 6.9:

$$L(E|p) = \binom{n}{x} p_i^x (1 - p_i)^{n-x}.$$

As an example, let us use the data in Example 6.7. We will use a discrete prior distribution, just as in Section 6.2.2.3. Unlike the examples in that earlier section, the present example uses an informed prior. Assume that a prior distribution was developed by plant equipment experts based on population variability data from similar systems, but adapted to account for untested new design aspects of this system. The prior is defined on 81 points, for p = 0, 0.01, 0.02, ..., 0.8. The most likely value is p = 0.1. From there, the prior falls linearly until p = 0.3, then tails off to 0 at p = 0.8. On the low end it falls linearly to 0 at p = 0. The distribution is shown in Figure 6.26.

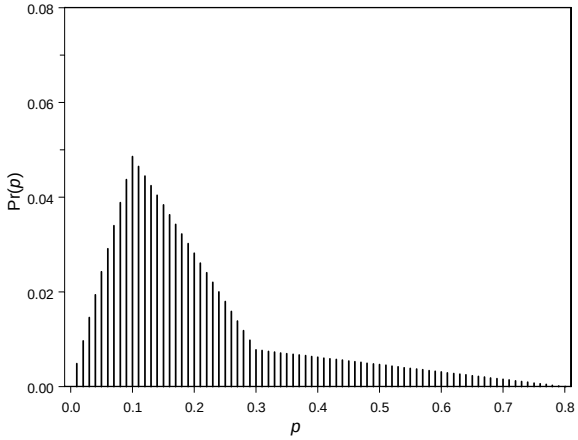


Figure 6.26 Discrete informative prior distribution for p.

This prior is discrete. We will want to compare the prior with the posterior distribution, and the graph of two discrete distributions is easier to read if just the tops of the vertical bars are plotted. The resulting plot of the prior looks like a continuous density, but it still is intended to represent the above discrete distribution.

The likelihood function is shown in Figure 6.25. The posterior distribution is proportional to the product of the prior and the likelihood, normalized so that the total probability equals 1. Figure 6.27 shows the prior and the posterior distributions on the same plot.

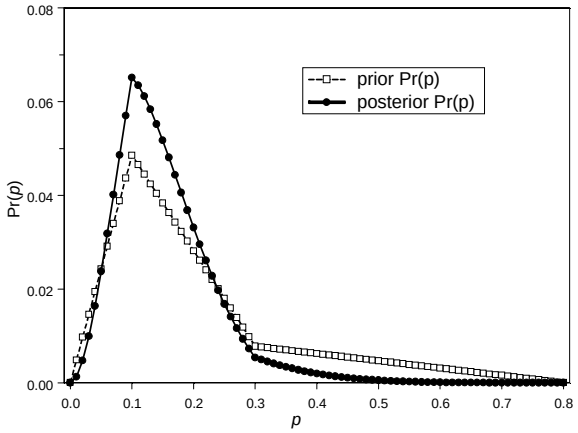


Figure 6.27 Discrete prior and posterior distribution for data in Example 6.7.

Note that the posterior follows the shape of the prior very closely. This is because the data are consistent with the peak area of the prior, but are not yet strong enough to appreciably reduce the uncertainty in the prior — there are only eight demands.

What happens to this posterior as additional data accumulate? Suppose that ten times as much data had been collected, 10 failures in 80 demands. The likelihood function, given by Equation 6.9 with this new data set, is shown in Figure 6.28.

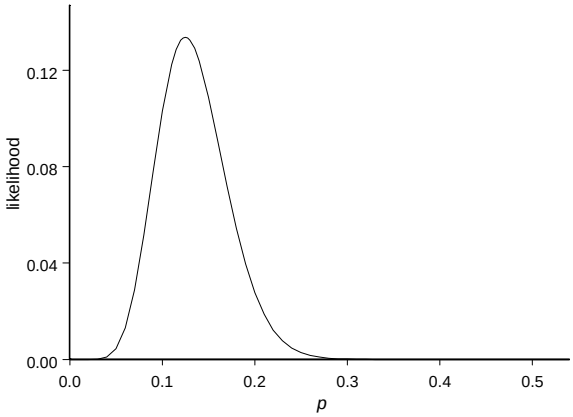


Figure 6.28 Likelihood as a function of p, for ten times the data of Example 6.7.

The posterior distribution is proportional to the product of the prior and this new likelihood. Figure 6.29 shows the prior and this new posterior.

Table 6.10 compares the results of the Bayesian analyses with the original data and with ten times as much data.

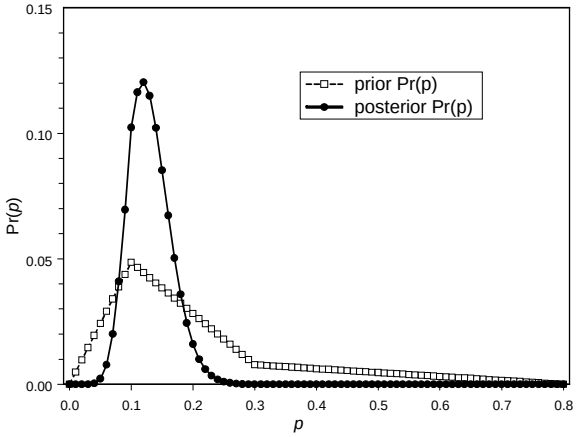


Figure 6.29 Discrete prior and posterior distributions for *p*, with 10 times as much data as in previous figure.

Table 6.10 Comparison of Bayesian distributions.

Distribution	5th %tile	mean	95th %tile
Prior	0.04	0.206	0.54
Posterior, original data	0.05	0.153	0.29
Posterior, ten times more confirmatory data	0.07	0.130	0.19

The difference between the two posterior distributions results from the differences between the two likelihoods. In this hypothetical example, both data sets have the same MLE, 0.125, but the larger data set has a likelihood that is more concentrated. The posterior distribution from the larger data set is dominated by the likelihood, and closely resembles it.

Readers are strongly encouraged to work through a few examples like this on their own. The calculations are easy to carry out with a spreadsheet.

6.3.2.2 Estimation with a Conjugate Prior

We now consider the use of continuous prior distributions, beginning with a very convenient family of distributions, the conjugate priors.

6.3.2.2.1 Definitions

By far the most convenient form for the prior distribution of p is a beta(α_{prior}, β_{prior}) distribution. The beta distributions are the conjugate family for binomial data. The properties of the beta distribution are therefore summarized here, as well as in Appendix A.7.8.

If p has a beta(α, β) distribution, the density is

$$f(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \quad .$$

For most applications the gamma functions in the front can be ignored — they only form a normalizing constant, to ensure that the density integrates to 1. The important feature of the density is that

$$f(p) \propto p^{\alpha-1}(1-p)^{\beta-1} \tag{6.11}$$

where the symbol $\propto$ denotes "is proportional to." The parameters of the distribution, α and β, must both be positive. The mean and variance of the distribution are

$$\mu = \alpha/(\alpha+\beta), \tag{6.12}$$

$$\text{variance} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \mu(1-\mu)/(\alpha+\beta+1). \tag{6.13}$$

The shape of the beta density depends on the size of the two parameters. If $\alpha < 1$, the exponent of p is negative in Equation 6.11, and therefore the density is unbounded as $p \to 0$. Likewise, if $\beta < 1$, the density is unbounded as $p \to 1$. If both $\alpha > 1$ and $\beta > 1$, the density is roughly bell shaped, with a single mode. Appendix A.7.8 shows graphs of some beta densities. Equation 6.13 shows that as the sum $\alpha + \beta$ becomes large, the variance becomes small, and the distribution becomes more tightly concentrated around the mean.

As will be seen below, if the prior distribution is a beta distribution, so is the posterior distribution. Therefore, the above statements apply to both the prior and the posterior distributions.

Appendix C tabulates selected percentiles of beta distributions. Also, the percentiles of a beta distribution can be found by many software packages, including some spreadsheets. Also, the percentiles can be obtained from algebraic formulas involving percentiles of the F distribution, as explained in Appendix A.7.8.

6.3.2.2.2 Update Formulas

The beta family is **conjugate** to binomial data. That is, updating a beta prior distribution with the data produces a posterior distribution that is also a beta distribution. This follows immediately from the derivation of the posterior distribution. By Bayes' theorem (Appendix B.5), the posterior distribution is related to the prior distribution by

$$f_{post}(p) \propto \Pr(X = x|\, p)\, f_{prior}(p)\,. \qquad (6.14)$$

This is the continuous analogue of Equation 6.10. As mentioned in the earlier sections, the probability of the data is also called the "likelihood." It is given by Equation 6.9. Stripped of all the normalizing constants, the beta p.d.f. is given by Equation 6.11.

Therefore, the beta distribution and the binomial likelihood combine as:

$$\begin{aligned} f_{post}(p) &\propto p^{x}(1-p)^{n-x}\, p^{\alpha-1}(1-p)^{\beta-1} \\ &\propto p^{x+\alpha-1}(1-p)^{n-x+\beta-1}\quad . \end{aligned}$$

In the final expression, everything that does not involve p has been absorbed into the proportionality constant. This shows that the posterior distribution is of the form beta(α_{post}, β_{post}), with

$$\begin{aligned} \alpha_{post} &= \alpha_{prior} + x \\ \beta_{post} &= \beta_{prior} + (n - x)\quad . \end{aligned}$$

The mean and variance of the prior and posterior distributions are given by Equations 6.12 and 6.13, using either the prior or posterior α and β.

These update formulas give intuitive meaning to the beta parameters: α_{prior} corresponds to a prior number of failures and β_{prior} to a prior number of successes. Assuming a beta(α_{prior}, β_{prior}) distribution is equivalent to having observed α_{prior} failures and β_{prior} successes before the current data were observed.

6.3.2.3 Possible Conjugate Priors

A concentrated distribution (small variance, large value of $\alpha_{prior} + \beta_{prior}$) represents much presumed prior knowledge. A diffuse prior (large variance, small value of $\alpha_{prior} + \beta_{prior}$) represents very little prior knowledge of p.

6.3.2.3.1 Informative Prior

The warning given in Section 6.2.2.5.1 applies here as well: the prior distribution must be based on information other than the data. If possible, relevant information from the industry should be used.

The calculations are now illustrated with Example 6.7, one failure to start in eight demands of the AFW turbine train. Poloski et al. (1998) examined nine years of data from many plants, and found a beta(4.2, 153.1) distribution for the probability of the AFW train failure to start.

Application of the update formulas yields

$$\begin{aligned} \alpha_{post} &= \alpha_{prior} + x = 4.2 + 1 = 5.2 \\ \beta_{post} &= \beta_{prior} + (n - x) = 153.1 + (8 - 1) = 160.1\,. \end{aligned}$$

The mean of this distribution is

5.2/(5.2 + 160.1) = 0.031,

and the variance is

0.031×(1 − 0.031)/(5.2 + 160.1 + 1) = 1.89E−4,

and the standard deviation is the square root of the variance, 0.014. The 5th and 95th percentiles of the posterior beta(α, β) distribution are found from Table C.5, except the tabulated β values do not go above 100. A footnote to that table gives an approximation that is valid for $\beta >> \alpha$. That formula applies, because 160.1 >> 5.2. According to the formula the q quantile is approximated by

$$\chi^2_q(2\times 5.2)/[2\times 160.1 + \chi^2_q(2\times 5.2)].$$

Therefore the 5th percentile of the beta distribution is approximately

$$\chi^2_{0.05}(10.4)/[320.2 + \chi^2_{0.05}(10.4)] = 4.19/[320.2 + 4.19] = 0.013,$$

and the 95th percentile is approximately

$$\chi^2_{0.95}(10.4)/[320.2 + \chi^2_{0.95}(10.4)] = 18.86/[320.2 + 18.86] = 0.056\,.$$

All these quantities are unitless.

The prior density, posterior density, and posterior c.d.f. of p are shown in Figures 6.30 through 6.32.

The posterior density is slightly to the right of the prior density. It is to the right because the data, one failure in eight demands, show worse performance than the industry history. The posterior density is only slightly different from the prior density because the data set is small compared to the industry experience (eight demands in the data and an effective 157.3 demands for the industry).

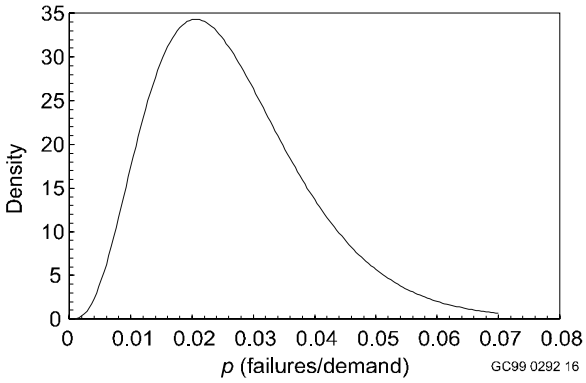


Figure 6.30 Prior density for p, beta(4.2, 153.1).

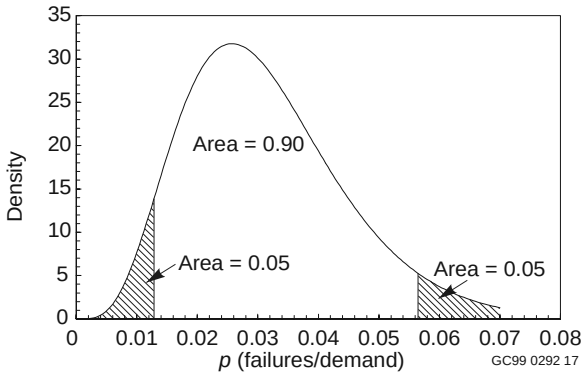


Figure 6.31 Posterior density for p, beta(5.2, 160.1). The 5th and 95th percentiles are shown.

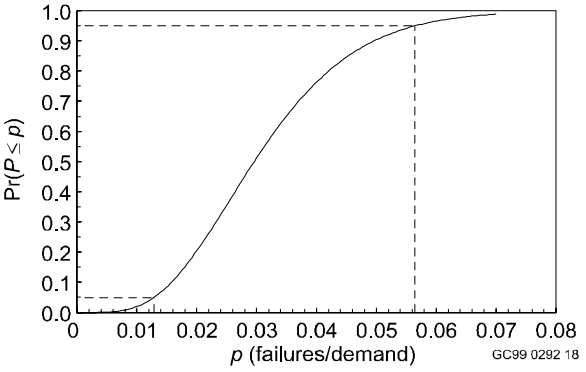


Figure 6.32 Posterior cumulative distribution of p. The 5th and 95th percentiles are shown.

The 5th and 95th percentiles are shown for the posterior distribution, both in the plot of the density and in the plot of the cumulative distribution.

6.3.2.3.2 Noninformative Prior

The Jeffreys noninformative prior is beta(½ , ½); see Box and Tiao (1973), Sections 1.3.4-1.3.5. This density is shown in Figure 6.33. It is not the uniform distribution, which is a beta(1, 1) distribution, but instead rises sharply at the two ends of the interval (0, 1).

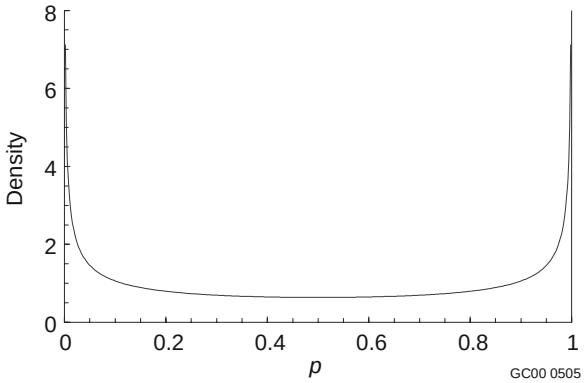


Figure 6.33 Jeffreys noninformative prior distribution for p.

Although the uniform distribution is sometimes used to model no prior information, there are theoretical reasons for preferring the Jeffreys noninformative prior. These reasons are given by Box and Tiao, and are suggested by the comparison with confidence intervals presented below. The uniform distribution would correspond intuitively to having seen one failure in two demands, which turns out to be too informative. The Jeffreys noninformative prior corresponds to having seen one-half a failure in one demand.

The Bayes posterior distribution for p, based on the Jeffreys noninformative prior, is beta($x + ½, n ! x + ½$). The mean of the distribution is $(x + ½)/(n + 1)$. Selected percentiles are tabulated in Appendix C.

The posterior distribution given here is very similar to the distributions used in the formulas for confidence intervals in Section 6.3.1.3. The only difference is in the parameters. The parameters here are averages of the parameters used in the confidence intervals. For example, the first parameter for the lower confidence limit is x, and the first parameter for the upper confidence limit is x+1. The Bayesian limits, on the other hand, use the same parameters for the entire posterior distribution, and the first parameter is $x + ½$, the average of the corresponding values for the confidence limits.

In Example 6.7, failure to start of the turbine-driven AFW train, the posterior distribution is beta(1.5, 7.5). The posterior mean is 1.5/(1.5 + 7.5) = 0.17. The posterior 90% interval is (0.023, 0.40). As is always the case with discrete data, the confidence interval is conservative, and so is wider than the Jeffreys credible interval. However, the two intervals are similar to each other, being neither to the right nor the left of the other. Tabular and graphical comparisons are given later.

6.3.2.3.3 Constrained Noninformative Prior

This prior distribution is a compromise between an informative prior and the Jeffreys noninformative prior. As was the case in Section 6.2.2.5.3, the prior mean, denoted here as p_0, is based on prior belief, but the dispersion is defined to correspond to little information. The priors are described by Atwood (1996) and by references given there.

For binomial data, the constrained noninformative prior distribution is not as neat as for Poisson data. The exact constrained noninformative prior has the form

$$f_{\text{prior}}(p) \propto e^{bp} p^{-1/2} (1 - p)^{-1/2}, \qquad (6.15)$$

where b is a number whose value depends on the assumed value of the mean, p_0. The parameter b is positive when $p_0 > 0.5$ and is negative when $p_0 < 0.5$. Thus, in typical PRA analysis b is negative. Atwood (1996) gives a table of values, a portion of which is reproduced in Appendix C as Table C.8. The table gives the parameter b of the distribution for selected values of p_0. In addition, it gives a beta distribution that has the same mean and variance as the constrained noninformative prior.

The beta approximation is illustrated here, and the exact constrained noninformative distribution is treated more fully in the section below on nonconjugate priors.

Return again to Example 6.7, the AFW turbine train failure to start. Let us use the mean of the industry prior found above, 4.2/157.3 = 0.0267. However, suppose that the full information for the industry prior is not available, or that the system under consideration is considered atypical so that the industry prior is not fully relevant. Therefore, the beta-approximation of the constrained noninformative prior will be used.

Interpolation of Table C.8 at $p_0 = 0.0267$ yields $\alpha = 0.4585$. Solving $\beta = \alpha(1 - p_0)/p_0$ gives $\beta = 16.7138$. The resulting posterior distribution has parameters 1.4585 and 23.7138. Interpolation of Table C.5 gives a 90% interval of (0.0068, 0.15).

6.3.2.3.4 Example Comparison of Above Methods

Just as in Section 6.2, the following general statements can be made:

- The Jeffreys noninformative prior results in a posterior credible interval that is numerically similar to a confidence interval.
- If the prior mean exists, the posterior mean is between the prior mean and the MLE.
- If two prior distributions have about the same mean, the more concentrated (less diffuse) prior distribution will yield the more concentrated posterior distribution, and will pull the posterior mean closer to the prior mean.

Figure 6.34 and Table 6.11 summarize the results of analyzing the AFW-failure-to-start data in the four ways given above.

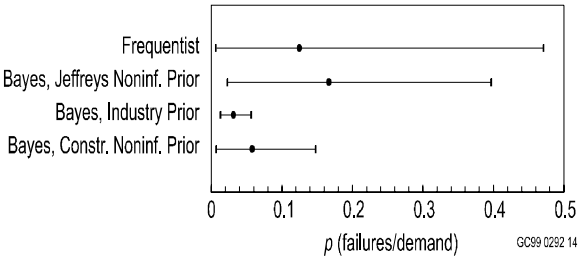


Figure 6.34 Comparison of four point estimates and interval estimates for p.

As in Section 6.2.2.5.4, the Jeffreys prior and the frequentist approach are listed next to each other because they give numerically similar results. The Jeffreys prior yields a posterior credible interval that is strictly contained in the confidence interval, neither to the right nor to the left.

In each Bayesian case, the posterior mean falls between the prior mean and the MLE, 0.125. The prior distribution has more influence when the prior distribution is more tightly concentrated around the mean. One measure of the concentration (at least when the means are similar) is the sum $\alpha_{\text{prior}} + \beta_{\text{prior}}$, because it corresponds to the total number of prior demands, and it is in the denominator of the variance in Equation 6.13. In the present example, when the prior distributions in Table 6.11 are ordered by increasing values of $\alpha_{\text{prior}} + \beta_{\text{prior}}$, the order is the noninformative prior, then the approximate constrained noninformative prior, and finally the industry prior. The three 90% intervals for the corresponding posterior distributions have decreasing length in the same order.

Table 6.11 Comparison of estimates with one failure in eight demands.

Method	Prior mean	Posterior parameters	Point estimate (MLE or posterior mean)	90% interval (confidence interval or posterior credible interval)
Frequentist	NA	NA	0.125	(0.0064, 0.47)
Bayes with Jeffreys noninformative prior, beta(0.5, 0.5)	0.5	$\alpha = 1.5$ $\beta = 7.5$	0.17	(0.022, 0.40)
Bayes with industry prior, beta(4.2, 153.1)	0.027	$\alpha = 5.2$ $\beta = 160.1$	0.031	(0.013, 0.056)
Bayes with approx. constrained noninform. prior, beta(0.4585, 16.7138)	0.027	$\alpha = 1.4585$ $\beta = 23.7138$	0.058	(0.0068, 0.15)

6.3.2.4 Estimation with a Continuous Nonconjugate Prior

Just as for β, continuous nonconjugate priors for p cannot be updated with simple algebra. The resulting posterior distribution does not have a simple form. Therefore, to enter it as the distribution of a basic event in a PRA code, a discrete approximation of the distribution must usually be used.

The posterior distribution must be obtained by numerical integration or by random sampling. Three methods are mentioned here, and the analyst may choose whatever seems easiest.

6.3.2.4.1 Direct Numerical Integration

To use numerical integration, use Equation 6.14 and write the posterior distribution as the product of the likelihood and the prior distribution:

$$Cf_{post}(p) = p^x(1 - p)^{n-x} f_{prior}(p) . \qquad (6.16)$$

Here C is a constant of proportionality. All the normalizing constants in f_{prior} and in the likelihood may be absorbed into C, leaving only the parts that depend on p on the right-hand side of the equation. Integrate $Cf_{post}(p)$ from 0 to 1. That is, integrate the right hand side of Equation 6.16. This integral equals C, because the integral of f_{post} must equal 1. Divide both sides of Equation 6.16 by the just-found constant C, to obtain the function f_{post}. Use numerical integration to find the moments and percentiles of this distribution. Some suggested methods of numerical integration are mentioned in Section 6.2.2.6.

6.3.2.4.2 Simple Random Sampling

To use random sampling, follow the rejection algorithm given in Section 6.2.2.6. The general algorithm, given in Section 6.2.2.6, can be restated for binomial data as follows. Define

$$m = (x/n)^x(1 - x/n)^{n-x}$$

if $0 < x < n$. If $x = 0$ or $x = n$, define $m = 1$. The steps of the algorithm are:

(1) Generate a random p from the prior distribution.
(2) Generate u from a uniform distribution, $0 \le u \le 1$.
(3) If $u \le p^x(1 - p)^{n-x}/m$, accept p in the sample. Otherwise discard p.

Repeat Steps (1) through (3) until a sample of the desired size is found.

6.3.2.4.3 More Complicated Random Sampling

All-purpose Bayesian update programs can be used for the present simple problem, just as in Section 6.2. The powerful program BUGS is mentioned in Section 6.2.2.6.3, and described more fully in Sections 7.2.3 and 8.3.3.3. It can be used here, although it is intended for much more complicated problems.

6.3.2.5 Examples with Nonconjugate Priors

Several possible nonconjugate prior distributions are discussed here.

6.3.2.5.1 Lognormal Distribution

The lognormal distribution is by far the most commonly used nonconjugate distribution. The parameter p has a lognormal distribution if $\ln(p)$ is normally distributed with some mean μ and variance σ^2.

Facts about the lognormal distribution are given in Appendix A.7.3. One important fact is that the range of the lognormal distribution is from 0 to $+\infty$. Thus, the distribution of p cannot be exactly lognormal, because p cannot be greater than 1. When using a lognormal prior, one must immediately calculate the prior $\Pr(p > 1)$. If this probability is very small, the error can be neglected. (When generating values p from the lognormal distribution, either throw away any values greater than 1 or set them equal to 1. In either case, such values hardly ever occur and do not affect the analysis greatly.) On the other hand, if the prior $\Pr(p > 1)$ is too large to be negligible, then the lognormal distribution cannot possibly be used. Even if the software accepts the lognormal distribution, and hides the problem by somehow handling the values that are greater than 1, the actual distribution used is not lognormal. It is truncated lognormal, or lognormal with a spike at 1, with a different mean and different percentiles from the initially input lognormal distribution. The analyst's two options are to recognize and account for this, or to use a different prior distribution.

To use the above sampling algorithm with a lognormal prior, p must be generated from a lognormal distribution. The easiest way to do this is first to generate z from a standard normal distribution, that is, a normal distribution with mean = 0 and variance = 1. Many software packages offer this option. Then, let $y = \mu + \sigma z$, so that y has been generated from a normal(μ, σ^2) distribution. Finally, let $p = e^y$. It follows that p has been randomly generated from the specified lognormal distribution.

6.3.2.5.2 Logistic-Normal Distribution

This distribution is explained in Appendix A.7.9. The parameter p has a logistic-normal distribution if $\ln[p/(1 - p)]$ is normally distributed with some mean μ and variance σ^2. The function $\ln[p/(1 - p)]$ is called the **logit** function of p. It is an analogue of the logarithm function for quantities that must lie between 0 and 1. Using this terminology, p has a logistic-normal distribution if $\text{logit}(p)$ is normally distributed.

Properties of the logistic-normal distribution are given in Appendix A.7.9, and summarized here. Let $y = \ln[p/(1 - p)]$. Then $p = e^y / (1 + e^y)$. This is the inverse of the logit function. As p increases from 0 to 1, y increases from $-\infty$ to $+\infty$.

Note, unlike a lognormally distributed p, a logistic-normally distributed p must be between 0 and 1. Therefore, the logistic-normal distribution could be used routinely by those who like the lognormal distribution, but do not know what to do when the lognormal distribution assigns p a value that is greater than 1.

The relation between p and $y = \text{logit}(p)$ gives a way to quantify prior belief about p in terms of a logistic-normal distribution. First, decide on two values, such as lower and upper plausible bounds on p or a median and plausible upper bound, equate them to percentiles of p, translate those percentiles to the corresponding two percentiles of the normal random variable Y, and finally, solve those two equations for μ and σ.

To generate a random value from a logistic-normal distribution, first generate y from a normal (μ, σ^2) distribution, exactly as in the section above on the lognormal distribution. Then let $p = e^y / (1 + e^y)$. This p has been randomly generated from the specified logistic-normal distribution.

6.3.2.5.3 Exact Constrained Noninformative Distribution

The prior distribution has the form of Equation 6.15, and the posterior distribution is

$$f_{\text{post}}(p) = C_1 e^{bp} p^{x - 1/2} (1 - p)^{n - x - 1/2},$$

where C_1 is a normalizing constant to make the density integrate to 1.0. Except for the normalizing constant, this is e^{bp} times a beta($x+½$, $n-x+½$) distribution. Numerical integration is straightforward, and will not be explained here. To generate a sample from the posterior distribution, the rejection method algorithm originally given in Section 6.2.2.6 takes the following form.

Write the beta($x+½$, $n-x+½$) density as

$$f_{\text{beta}}(p) = C_2 p^{x - 1/2} (1 - p)^{n - x - 1/2}.$$

Typically, the desired mean of p is less than 0.5; if it is not, reverse the roles of p and $1 - p$. The algorithm first defines M to be the maximum possible value of the ratio $f_{\text{post}}(p) / f_{\text{beta}}(p)$. Because $b < 0$ in Table C.8, we have $e^{bp} \le 1$, making M equal to C_1/C_2. Therefore, the condition in Step (3) of the algorithm reduces to

$$u \le e^{bp}.$$

Therefore, the algorithm simplifies to the following:

(1) Generate a random p from the beta(x+½, n ! x+½) distribution. Ways to do this are discussed below.
(2) Generate u from a uniform distribution, 0 # u # 1.
(3) If u # e^{bp}, accept p in the sample. Otherwise discard p.

Repeat Steps (1) through (3) until a sample of the desired size is found.

Not all standard software packages give the option of generating random numbers from a beta distribution, although many more allow random number generation from a gamma distribution or from a chi squared distribution. When working with such software, let y_1 be randomly generated from a gamma(x+½, 1) distribution and let y_2 be randomly generated from a gamma(n ! x+½, 1) distribution. Alternatively, let y_1 be randomly generated from a chi-squared(2x+1) distribution and let y_2 be randomly generated from a chi-squared(2n ! 2x+1) distribution. In either case, define p = $y_1/(y_1+y_2)$. Then, p has been generated from the specified beta(x+½, n ! x+½) distribution. (See Chapter 25 of Johnson et al. 1995.)

6.3.2.5.4 Maximum Entropy Prior

The maximum entropy prior and the constrained noninformative prior were developed with the same goal: to produce a diffuse distribution with a specified plausible mean. The diffuseness of the maximum entropy distribution is obtained by maximizing the entropy, defined as

$$-E[\ln f(p)] = -\int [\ln f(p)] f(p) dp .$$

When p is restricted to the range from 0 to 1, it can be shown that the density f maximizing the entropy is uniform,

$f(p) = 1$ for 0 # p # 1

and $f(p)$ = 0 elsewhere. More interesting is the case when the mean of the distribution is required to equal some prespecified value p_0. In this case the maximum entropy distribution has the form of a truncated exponential distribution,

$f(p) = Ce^{bp}$ for 0 # p # 1

and $f(p)$ = 0 elsewhere. In this form, b is negative when p_0 < 0.5 and b is positive when p_0 > 0.5. The value of b corresponding to a particular mean must be found by numerical iteration. Some authors write $e^{!bp}$ instead of e^{bp}; this simply reverses the sign of the parameter b.

The maximum entropy distribution and the uniform distribution are related — if the constraint on the mean is removed, the maximum entropy distribution equals the uniform distribution. In this sense, the maximum entropy distribution is a generalization of the uniform distribution. The constrained noninformative distribution is the same sort of generalization of the Jeffreys noninformative distribution — if the constraint is removed, the constrained noninformative prior becomes the Jeffreys noninformative prior. Atwood (1996) reviews the reasons why the Jeffreys prior is superior to the uniform prior, and uses the same reasoning to argue that the constrained noninformative prior is superior to the maximum entropy prior.

In practice, it may make little difference which distribution is used. Both distributions are intended to be used when little prior knowledge is available, and quantifying "little prior knowledge" is not something that can be done precisely.

Sampling from the posterior distribution is similar to the other sampling procedures given above, so most of the details are not given. The only point deserving discussion is how to generate a random sample from the maximum entropy prior. The most convenient method is the **inverse c.d.f. algorithm**. This algorithm is simple in cases when the c.d.f. and its inverse can be calculated easily.

For example, let the random variable P have c.d.f. F. Let $F^{!1}$ be the inverse function, defined by $u = F(p)$ if and only if $p = F^{!1}(u)$. Let U be defined as $F(P)$. What is the distribution of U? The c.d.f. of U is found by

$$\begin{aligned} \Pr(U \# u) &= \Pr[\, F(P) \# u \,] \\ &= \Pr[\, P \# F^{!1}(u) \,] \\ &= F[\, F^{!1}(u) \,] \quad \text{because } F \text{ is the c.d.f. of } P \\ &= u . \end{aligned}$$

Therefore, U has a uniform distribution. The letter U was not chosen by accident, but in anticipation of the uniform distribution.

To generate a random value p from the distribution F, generate a random u from the uniform (0, 1) distribution, something that many software packages allow. Then define $p = F^{!1}(u)$. This is the inverse c.d.f. method of random number generation.

To apply this to the maximum entropy distribution, first integrate the maximum entropy density to yield the c.d.f.

$F(p) = (1 ! e^{bp})/(1 ! e^{b})$.

Generate u from a uniform(0, 1) distribution, and set

$$u = (1 - e^{bp})/(1 - e^{b}) .$$

Solve this equation for p,

$$p = -\ln[\, 1 - (1 - e^{b})u \,]/b .$$

Then, p has been randomly generated from the maximum entropy distribution. Repeat this with new values of u until enough values of p have been obtained.

6.3.2.5.5 Example Calculation

These techniques will be illustrated with the Example 6.7, one failure to start in eight demands of the AFW turbine train. Two prior distributions will be assumed, the lognormal prior used by the Accident Sequence Evaluation Program (ASEP), as presented by Drouin et al. (1987), and a logistic-normal distribution having the same 50th and 95th percentiles.

The ASEP distribution for turbine-driven pump failure to start is lognormal with mean 3E-2 per demand and error factor 10. The three relevant equations from Appendix A.7.3 are

$$\mathrm{EF}(p) = \exp(1.645\sigma)$$
$$\mathrm{mean}(p) = \exp(\mu + \sigma^2/2)$$
$$p_q = \exp(\mu + \sigma z_q)$$

where the subscript q denotes the qth quantile, and z_q is the qth quantile of the standard normal distribution.

Solving the first equation yields $\sigma = 1.3997$. Substitution of this into the second equation yields $\mu = -4.4862$.

The percentiles are not needed yet, but the third equation gives the median, $p_{0.50} = \exp(\mu) = 0.01126$, and the 95th percentile, $p_{0.95} = \exp(\mu + 1.645\sigma) = 0.1126$. (The relation of these two percentiles can also be derived from the fact that the error factor equals 10.)

The prior $\Pr(p > 1)$ is 6.75E-4, a very small number. In the calculations of this section, the lognormal distribution is truncated at 1.0. That is, integrals are renormalized to make the integral of the density from 0 to 1 equal to exactly 1.0. If random sampling is performed, any sampled values that are greater than 1 are discarded.

The prior and posterior densities of p are shown in Figure 6.35. The densities were calculated using software for numerical integration.

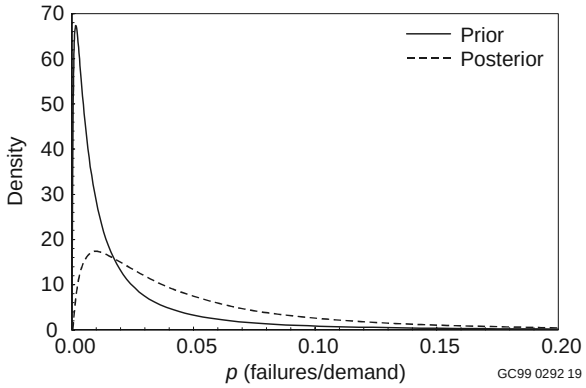


Figure 6.35 Lognormal prior density and posterior density for p.

As a second example, consider the logistic-normal prior distribution having the same 50th and 95th percentiles as the above lognormal prior. These percentiles are 0.01126 and 0.1126. To find the parameters of the underlying normal distribution, set $Y = \ln[p/(1 - p)]$. By the properties of the logistic-normal distribution given in Appendix A.7.9, the 50th and 95th percentiles of Y are

$$y_{0.50} = \ln[0.01126/(1 - 0.01126)] = -4.475$$
$$y_{0.95} = \ln[0.1126/(1 - 0.1126)] = -2.064 .$$

Because Y has a normal(μ, σ^2) distribution, it follows that

$$\mu = -4.475$$
$$\mu + 1.645\sigma = -2.064$$

so $\sigma = 1.466$.

Monte Carlo simulation shows that the truncated-lognormal and logistic-normal prior densities are virtually the same, with means, medians, 5th and 95th percentiles agreeing to two significant digits. As a consequence, the posterior distributions from the two priors are also nearly the same, although the means and percentiles may differ slightly in the second significant digit.

Numerical integration was used, but BUGS could have been used. As an illustration, the script for using BUGS is given in Figure 6.36.

This script assigns a logistic-normal prior distribution to p. If a lognormal prior is used instead, BUGS returns an error message during the simulation, presumably because it has generated a value of p greater than 1. The script assigns Y a normal distribution with mean -4.475. The second parameter is $1/\sigma^2$, because that is how BUGS parameterizes a normal distribution. The entered value, 0.4653,

equals $1/1.466^2$. The script then gives X a binomial(8, p) distribution. Finally, the line beginning "list" contains the data, the single observed value 1 in this example. BUGS also wants an initial value for p, but it is willing to generate it randomly.

```
model
{
  y ~ dnorm(-4.475, 0.4653)
  p <- exp(y)/( 1 + exp(y))
  x ~ dbin(p, 8)
}
list(x = 1)
```

Figure 6.36 Script for analyzing Example 6.7 with BUGS.

For the present example, the difference between the lognormal and logistic-normal priors is very small, having no effect on the posterior. The difference between the two priors can be important if the probability of failure is larger and/or the uncertainty is larger. That can be the case with some human errors, with hardware failures in unusually stressful situations, and with recovery from failure if recovery is modeled as an event separate from the original failure. For example, the NUREG 1150 PRA for Surry (Bertucio and Julius 1990) uses the lognormal distribution for most failure probabilities. However, some failure probabilities are large, considerably larger than 3E!2. In nearly all of those cases, the PRA does not use a lognormal distribution. Instead, the maximum entropy distribution is the PRA's distribution of choice. Other possible distributions, which were not widely known in the PRA community in 1990, would be the constrained noninformative distribution or a logistic-normal distribution.

6.3.2.6 Estimation with Fixed Number of Failures and Random Number of Demands

Sometimes it is useful to consider a random number of demands, a **waiting time**, to achieve a fixed number of failures x. For example, if the failure probability p is believed to change over time, only the most recent history may represent current behavior. In such a situation, one might decide to use only the most recent few failures, such as $x = 3$, and to treat the corresponding number of demands n as random. Here n is the number of demands counted backwards from the present to the xth failure in the past. Earlier failures and demands could be used to construct a prior distribution, but the dispersion of the prior distribution should be set large because the earlier events are not considered fully relevant to the present.

With such waiting-time data, the likelihood is proportional to

$$p^x(1 ! p)^{n ! x} .$$

Therefore, except for the normalizing constant the likelihood is the same as for binomial data. Therefore, it works out that Bayesian analysis with an informative prior is exactly the same whether the data are regarded as coming from a random count of x failures in a fixed number of demands, n, or a random number of demands, n, for a fixed number of failures, x. The posterior distributions are identical.

6.3.3 Model Validation

All the methods in this section are analogues of methods considered for failure rates, but the details are somewhat different. Some repetition is inevitable, but the examples in this section are chosen to complement the examples of Section 6.2.3, not to duplicate them. For a more complete appreciation of the model validation techniques, both this section and Section 6.2.3 should be read.

The comments at the start of Section 6.2.3 apply equally to this section, and must not be ignored. In particular, an analyst who estimates parameters should check the assumptions of the model. However, this section contains more than will be needed on any one analysis. The methods here are offered for possible use, and the analyst should select the appropriate ones.

The first assumption of the binomial model, given in Section 2.3.2, is that the probability of failure is the same on any demand. This assumption will be examined against two possible alternative assumptions: (1) different subsets of the data have different values of p, but in no special order; and (2) a time trend exists. The second assumption of the binomial model is that the outcome on one demand is statistically independent of the outcome on a different demand. This will be examined against the alternatives of common-cause failures and of clustering in time of the failures. Finally, the consistency of the prior distribution and the data will be considered.

One need not worry about whether n is really constant. If n is not constant, we may treat it as constant by conditioning on n, as explained in Section 2.3.2.4.2.

6.3.3.1 Poolability of Data Sources

The methods will be illustrated by data from diesel generator failures to start, shown in Example 6.8.

Example 6.8 EDG failures to start on demand.

Emergency diesel generator (EDG) failures to start on demand were recorded for three kinds of demands: unplanned demands, the tests performed once per operating cycle (approximately every 18 months), and the monthly tests. The counts are given below.

Type of demand	Failures to start	Number of demands
Unplanned	2	181
Cyclic test	17	1364
Monthly test	56	15000

Table C.1 of Grant et al. (1996) gives the data for the first two rows, at plants reporting under Regulatory Guide RG-1.108 during 1987-1993. The failures were those reported in LERs. The number of failures on monthly tests at those plants comes from the unpublished database used for that report, and the number of monthly demands was estimated in a very crude way for use in this example.

6.3.3.1.1 Graphical Technique

To explore the relations between subsets of the data, mark the subsets on one axis. For each of these subsets of the data, plot an estimate of *p* and a confidence interval for *p* against the other axis. Patterns such as trends, outliers, or large scatter are then visible.

In Example 6.8, the subsets are types of demand. The data set from each demand type is analyzed separately, and the graph shows an estimate and a confidence interval for each year, plotted side by side. This is shown in Figure 6.37. The plot was produced with a graphics package, although a hand-drawn plot would be adequate to show the results.

The plot shows that the unplanned demands and the cyclic tests appear to have similar values of *p*, but the monthly tests appear to have a lower value. Several reasons for the difference could be conjectured: the monthly tests may be less stressful, the failures may not all be reported in LERs, or the estimated number of demands may be badly incorrect.

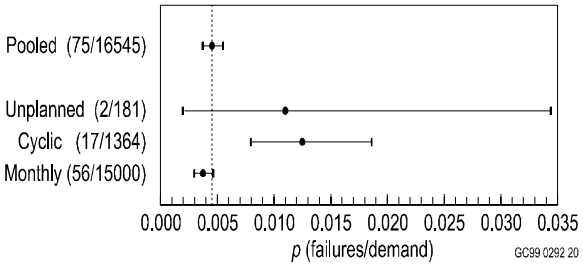


Figure 6.37 MLEs and 90% confidence intervals for *p*, for three types of demand and for the pooled data.

Figure 6.21, which is the corresponding plot in Section 6.2.3.1.1, has the cells (plants, in that example) arranged in order of decreasing $\hat{\lambda}$. Figure 6.37 does not order the cells by decreasing $\hat{p}$, because the number of cells is small, only three, and because the cells already have a natural order. The analyst must decide what order makes the most sense and is easiest for the user to interpret.

The interval for the pooled data is also shown, not because the data justify pooling, but simply as a reference for comparison. A dotted reference line is drawn through the point estimate based on the pooled data. If only a few data subsets need to be compared, as in Figure 6.37, these embellishments are unnecessary. With many subsets, however, the eye tends to get lost without the reference line. The reference line has the added advantage of focusing the eye on the confidence intervals rather than the point estimates.

The graph is only a picture. Pictures like these are useful, but cannot always be used in an easy way to draw conclusions about differences between data subsets. The warnings given in Section 6.2.3.1.1 deserve repetition:

- If many confidence intervals are plotted, all based on data with the same *p*, a few will be far from the others because of randomness alone. An outlying interval does not prove that the *p*s are unequal.
- This same statement is true if other intervals are used, such as Bayes credible intervals based on the noninformative prior. The issue is the random variability of data, not the kind of interval constructed.
- If there are few intervals, on the other hand, intervals that just barely overlap can give strong evidence for a difference in the *p*s.

To quantify the strength of the evidence seen in the picture, a formal statistical procedure is given in the next subsection. The picture gives a preview, and helps in the interpretation of the formal statistical quantification. In the present example, if the statistical test finds a statistically significant difference between data subsets, it is natural to then ask what kind of difference exists. The picture shows that p seems to be similar for the unplanned demands and for the cyclic tests, but smaller for the monthly tests. In this way, the picture provides insight, even though it does not provide a quantitative statistical test.

6.3.3.1.2 Statistical Tests

Simple Contingency Tables ($2 \times J$). The natural format for the data is a "contingency table." An introductory reference to this subject is Everitt (1992), and many general statistics texts also have a chapter on the topic. In a two-way table, two attributes of the events are used to define rows and columns, and the numbers in the table are counts. In the present example, two attributes of any event are the type of demand and whether it is a failure or success. One way to build a contingency table is to let the first row show system failures and the second row system successes. Then let the columns correspond to the demand types. (Of course, the roles of rows and columns can be reversed if that fits better on the sheet of paper.) The table entries are the counts of the events for each cell, shown in Table 6.12 for Example 6.8.

Table 6.12 Contingency table for Example 6.8.

	Unplanned	Cyclic	Monthly	Total
Failure	2	17	56	75
Success	179	1347	14944	16470
Total	181	1364	15000	16545

The essence of this table is a 2×3 table, because the basic data counts occupy two rows and three columns.

The row totals, column totals, and grand total are shown in the right and bottom margins. A general, two-way contingency table has I rows and J columns. (Although this discussion considers only $2 \times J$ tables, it does no harm to give the general formulas, keeping in mind that the examples of this section have $I = 2$.) The count in the ith row and jth column is denoted n_{ij}, for i any number from 1 to I and j from 1 to J. The total count in row i is denoted n_{i+} and the total count in column j is denoted n_{+j}. The grand total is denoted n_{++}.

For example, Table 6.12 has $n_{1,3} = 56$ and $n_{2,1} = 179$. It has $n_{2+} = 16470$ and $n_{+2} = 1364$. The grand total, n_{++}, equals 16545 in the example.

Let the null hypothesis be

H_0: p is the same for all the data subsets.

The alternative hypothesis is

H_1: p is not the same for all the data subsets.

In the example, the data subsets are the three demand types. The analyst must investigate whether H_0 is true. The method used is to see what kind of data would be expected when p really is the same, and then to see how much the observed counts differ from the expected. If the differences are small, the counts are consistent with the hypothesis H_0. If, instead, the differences are large, the counts show strong evidence against H_0.

If H_0 is true, that is, if p is really the same for all the demand types, the natural estimate of p is

$$\hat{p} = n_{1+} / n_{++} .$$

Then for column j, one would have expected $n_{+j}\hat{p}$ failures on average. This reasoning leads to the formula for the expected count in cell ij:

$$e_{ij} = n_{i+}n_{+j} / n_{++}.$$

In Table 6.12, for unplanned demands one would have expected 181×(75/16545) = 0.82 failures on average, for cyclic tests 1364×(75/16545) = 6.19 failures, and so forth.

The difference between the observed count and the expected count for any cell is $n_{ij} - e_{ij}$. There are many cells, and therefore many ways of combining the differences to yield an overall number. One useful way is to construct

$$X^2 = \sum_i \sum_j (n_{ij} - e_{ij})^2 / e_{ij}.$$

X^2 is called the chi-squared statistic, or sometimes the Pearson chi-squared statistic. Note, X^2 as defined here is slightly different from the chi-squared statistic for constant event rate in Section 6.2.3.1.2. In that section, the cells had one index, whereas in this section, the cells have two indices, and the expected counts are calcu-

lated differently. Other than that, the statistics are the same. Table 6.13 expands Table 6.12 to show the quantities needed to calculate X^2. The observed counts and the expected counts have the same totals, except for roundoff.

Table 6.13 Counts, expected counts, and contributions to X^2, for Example 6.8.

	Unplanned	Cyclic	Monthly	Total
Failure	2 0.82 1.70	17 6.19 18.92	56 68.00 2.12	75
Success	179 180.18 0.01	1347 1357.80 0.09	14944 14932 0.01	16470
Total	181	1364	15000	16545

For example, there were 2 failures on unplanned demands. The expected number of failures on unplanned demands, if H_0 is true, is 181×75/16545 = 0.82. And the contribution of that cell to X^2 is $(2 - 0.82)^2/0.82 = 1.70$.

When H_0 is true and the total count is large, the distribution of X^2 has a distribution that is approximately chi-squared with $(I\text{-}1)\times(J\text{-}1)$ degrees of freedom. In Table 6.12, the number of degrees of freedom is $(2-1)\times(3-1) = 2$. If X^2 is large compared to the chi-squared distribution, the evidence is strong that H_0 is false; the larger X^2, the stronger the evidence.

Interpretation of Test Results. Based on any 2×3 contingency table, such as Table 6.12, suppose that X^2 were 6.4. A table of the chi-squared distribution shows that 5.991 is the 95th percentile of the chi-squared distribution with 2 degrees of freedom, and 7.378 is the 97.5th percentile. After comparing X^2 to these values, an analyst would conclude that the evidence is strong against H_0, but not overwhelming. Quantitatively, the analyst would "reject H_0 at the 5% significance level, but not at the 2.5% significance level." This is sometimes phrased as "the p-value is between 0.05 and 0.025." See the bulleted list in Section 6.2.3.1.2, in the interpretation following Table 6.6, for other phrases that are sometimes used.

If instead X^2 were 1.5, it would lie between the 50th and the 60th percentiles of the chi-squared distribution, and therefore would be in the range of values that would be expected under H_0. The analyst could say "the observed counts are consistent with the hypothesis H_0," or "H_0 cannot be rejected," or "the evidence against H_0 is very weak." The analyst would not conclude that H_0 is true, because it probably is not exactly true to the tenth decimal place, but would conclude that it cannot be rejected by the data.

In fact, in Example 6.8 X^2 equals 22.8, as found by totaling the six contributions in Table 6.13. This number is far beyond the 99.5th percentile of the chi-squared distribution, so the evidence is overwhelming against H_0. Such an analysis contributed to the decision of Grant et al. (1999b) not to consider monthly tests in their report.

This example was chosen to illustrate that subsets of the data can correspond not only to different locations or different hardware (for example, different plants or systems), but also to different conditions, in this case different types of demands. In reality, the data analyst should consider various kinds of subsets; in this example, with data coming from many plants, the analyst should consider possible between-plant differences. The plots and chi-squared tests are exactly the same as given above.

This brings up a difficulty with the present example that has been carefully hidden until now. The hypothesis H_0 is that all the subsets of the data have the same p. A hidden hypothesis, never even proposed for testing, is that within each data subset, every demand has the same p. In fact, this turns out not to be the case. Based on only the unplanned demands and cyclic tests, Grant et al. (1999b) report that the difference between plants is statistically significant — the evidence is strong that p differs from plant to plant. This means that the above analysis must be refined to account for possible differences between plants. Such variation is discussed in Chapter 8 of this handbook.

Thus, the data set has two sources of variation, differences between demand types and also differences between plants. In such a situation, consideration of only one variable at a time can throw off the results if the data set is "unbalanced," for example, if the worst few plants also happen to have the most unplanned demands and the fewest monthly demands. If such between-plant differences are contaminating the EDG data in Example 6.8, the observed difference might not reflect anything about the nature of the demands, but only that the plants with EDG problems were underrepresented on the monthly tests. Example 6.9 shows hypothetical data under such a scenario.

If only the good plants are considered, or if only the bad plants are considered, the data of Example 6.9

show no difference between unplanned demands and tests. The estimated p is the same for unplanned demands and for tests, 0.2 from the bad plants' data and 0.02 from the good plants' data. However, if the data from good plants and bad plants are combined, the unplanned demands appear to have a much higher failure probability than do the tests, 0.07 versus 0.03. This erroneous conclusion is a result of ignoring differences in the data, the existence of two kinds of plants, when the data are unbalanced because the bad plants have a much higher percentage of unplanned demands. Such a situation is known as **Simpson's paradox**.

Example 6.9 Hypothetical unbalanced data.

Suppose that the industry consists of "bad" plants and "good" plants. The bad plants have a relatively high probability of failure to start, and also have relatively many unplanned demands. Suppose that the tests perfectly mimic unplanned demands, so that at either kind of plant p is the same on an unplanned demand and on a test. Data from such an industry might be given in the table below. The tables entries show failures/demands.

	Unplanned	Tests
Bad plants	4/20 = 0.2	4/20 = 0.2
Good plants	1/50 = 0.02	8/400 = 0.02
Totals	5/70 = 0.07	12/420 = 0.03

In fact, this scenario cannot be greatly influencing the data in Example 6.8, because most of the demands are periodic. Therefore, every plant must have approximately the same fraction of monthly tests and of cyclic tests. In conclusion, although between-plant variation must be considered, it is hard to imagine that it affects the outcome in Example 6.8.

As mentioned in Section 6.2.3.1.2, a full data analysis must not stop with the calculation of a p-value. In the present example, with a very large number of demands, it may be that the statistically significant difference is not very important from an engineering viewpoint. In other words, a large data set can detect differences in the second decimal place — differences that are not worth worrying about in practice.

This concern is addressed in the example by Figure 6.37, which shows that the probability of FTS is about 1/3 as large on monthly tests as on other demands, at least according to the reported data. Therefore, the difference is substantial in engineering terms, and the engineering portion of the data analysis can investigate reasons for the difference.

Required Sample Size. The above approach is valid if the values of n_{ij} are "large." If they are small, X^2 has a discrete distribution, and so cannot have a chi-squared distribution. As a rather extreme example, if n_{++}, the total number of demands, were equal to four in the framework of Example 6.8, there would only be a few ways that the four demands (and the number of failures, at least zero and at most four) could be arranged among the three demand types. Therefore X^2 could only take a few possible values.

Therefore, the user must ask how large a count is necessary for the chi-squared approximation to be adequate. An overly conservative rule is that all the expected cell counts, e_{ij}, be 5.0 or larger. Despite its conservatism, this rule is still widely used, and cited in the outputs of some current statistics packages. For a $2\times J$ table, Everitt (1992, Sec. 3.3), citing work by Lewontin and Felsenstein (1965), states that the chi-squared approximation is adequate if all the values of e_{ij} are 1.0 or greater, and that in "the majority of cases" it is sufficient for the e_{ij} values to be 0.5 or greater. For a 2×2 table, however, it is generally best not to use the chi-squared approximation at all, but to use the p-value from "Fisher's exact two-sided test," discussed below.

If the expected cell counts are so small that the chi-squared approximation appears untrustworthy, the analyst has two choices: (a) Pool some columns, thereby combining cells and increasing the expected cell counts. For example, in an investigation of differences between years, with few failures, it might be necessary to combine adjacent years so that the expected number of failures in each time-bin is at least 0.5; or (b) Some statistical software packages can compute the "exact distribution" of X^2 in some cases (typically for small tables). Conditional on the n_{i+} values and n_{+j} values, this exact distribution is the finite set of values that X^2 can possibly take, together with their associated probabilities. If the analyst is willing to base the decision on this conditional distribution, the exact distribution can be used. The commercial package StatXact performs such calculations using modern, fast algorithms, even for large tables, subject only to the memory available in the machine. In the special case of a 2×2 contingency table, many software packages compute this p-value, calling it the p-value from "Fisher's exact two-sided test." In general, the p-value from Fisher's exact test is preferable to the p-value

from the chi-squared approximation, and should be used whenever the software produces it. This, and other considerations for a 2×2 table, are discussed by Everitt (1992) and Atwood (1994).

In Table 6.13, the smallest expected count is e_{11} = 0.82. All the other expected counts are larger than 1.0. This indicates that the sample size is large enough.

6.3.3.2 No Time Trend

This section uses the unplanned HPCI demands from Example 6.5, with the failures indicated. To make a data set with a moderate number of failures, all types of failures are counted together, including failure to start, failure to run, failure of the injection valve to reopen after operating successfully earlier in the mission, and unavailability because of maintenance. For the example, no credit is taken for failures that were recovered. The data are given as Example 6.10.

Example 6.10 Dates of HPCI failures and unplanned demands, 1987-1993.

The HPCI demands of Example 6.5 are listed here with an asterisk marking demands on which some kind of failure occurred. The demands dates are given in columns, in format MM/DD/YY.

01/05/87*	08/03/87*	03/05/89	08/16/90*	08/25/91
01/07/87	08/16/87	03/25/89	08/19/90	09/11/91
01/26/87	08/29/87	08/26/89	09/02/90	12/17/91
02/18/87	01/10/88	09/03/89	09/27/90	02/02/92
02/24/87	04/30/88	11/05/89*	10/12/90	06/25/92
03/11/87*	05/27/88	11/25/89	10/17/90	08/27/92
04/03/87	08/05/88	12/20/89	11/26/90	09/30/92
04/16/87	08/25/88	01/12/90*	01/18/91*	10/15/92
04/22/87	08/26/88	01/28/90	01/25/91	11/18/92
07/23/87	09/04/88*	03/19/90*	02/27/91	04/20/93
07/26/87	11/01/88	03/19/90	04/23/91	07/30/93
07/30/87	11/16/88*	06/20/90	07/18/91*	
08/03/87*	12/17/88	07/27/90	07/31/91	

6.3.3.2.1 Graphical Techniques

Just as elsewhere in this chapter, the time axis can be divided into bins, and the data can be analyzed separately for each bin and compared graphically.

For Example 6.10, defining the bins to be years leads to Table 6.14. This leads to a plot similar to Figures 6.21 and 6.22, shown in Figure 6.38. The plot with the example data shows no evidence of a trend.

Table 6.14 HPCI failures on demand, by year.

Calendar year	Failures	Demands
1987	4	16
1988	2	10
1989	1	7
1990	3	13
1991	2	9
1992	0	6
1993	0	2

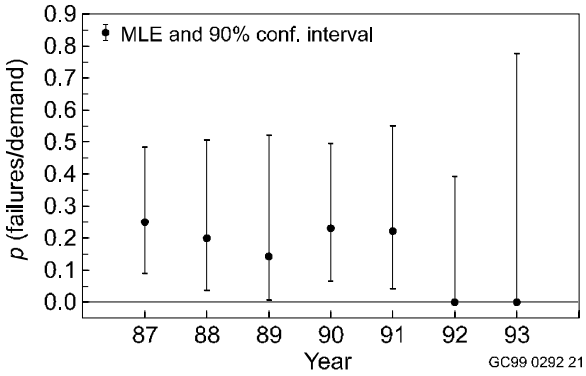


Figure 6.38 Point and interval estimates of p, each based on one year's data.

A plot that does not require a choice of how to construct bins is given in Figure 6.39, the analogue of Figure 6.23. It can be constructed when the demands can be ordered sequentially, as is the case for Example 6.10. In this plot, the cumulative number of failures is plotted against the cumulative number of demands. To help the eye judge curvature, a straight line is drawn, connecting the origin with the dot at the upper right.

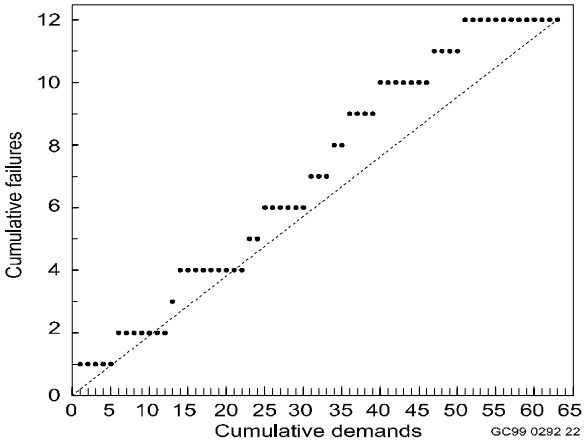


Figure 6.39 Cumulative number of failures versus cumulative number of demands.

The **slope** of any part of the graph is the vertical distance divided by the horizontal distance, $\Delta y / \Delta x$. In the present figure the horizontal distance is the number of demands that have occurred, and the vertical distance is the corresponding number of failures. Therefore,

slope = (number of failures)/(number of demands) ,

so the slope is a visual estimator of p. A roughly constant slope, that is, a roughly straight line, indicates a constant p. A changing slope indicates changes in p.

In Figure 6.39, the slope is relatively constant, indicating that p does not seem to change with time. This agrees with Figure 6.38. It is not clear whether the slight departure from the diagonal line in the right half of the figure is more than can be attributed to random variation. Such questions must be addressed by statistical tests, given below.

The details of the diagonal line probably do not matter. The line shown is the maximum likelihood estimate of the expected height of the plot at any horizontal point, assuming constant p. Other lines, slightly different, could also be justified.

6.3.3.2.2 Statistical Tests for a Trend in *p*

In this section, the null hypothesis remains

H_0: p is the same for all the data subsets.

but the alternative is now

H_1: p is either increasing or decreasing over time .

The Chi-Squared Test. This is the same test as given in Section 6.3.3.1.2, except now the data subsets are years or similar bins of time.

The data of Table 6.14 can be written as a 2×7 contingency table. The smallest expected cell count corresponds to failures in 1993, with the expected count = 2×12/63 = 0.4. This is too small to justify calculating a p-value from the chi-squared distribution. The problem can be remedied by pooling the two adjacent years with the smallest number of demands, 1992 and 1993. (Note, the decision of which subsets to pool is based on the numbers of demands only, not on whether or not those demands resulted in failures. Pooling based on demand counts is legitimate. Pooling based on the failure counts is not.)

When this 2 × 6 contingency is analyzed by the chi-squared test, the p-value is 0.77, indicating no evidence at all of differences between years. This is no surprise.

The Wilcoxon-Mann-Whitney Test. This test is similar in spirit to the Laplace test for a trend in λ. The null hypothesis is that p is the same for all demands. Suppose that the individual demands are in a known sequence. Against the alternative hypothesis that the failures tend to occur more at one end of the sequence — that is, p is either an increasing or a decreasing function of the sequence number — use the Wilcoxon-Mann-Whitney test, described in texts that cover nonparametric statistics. Two good sources of standard nonparametric methods are Conover (1999), and Hollander and Wolfe (1999). Hollander and Wolfe call this test the **Wilcoxon rank sum test**.

The test is based on the sum of the ranks of the failures. For example, in the sequence of failures and successes

failure, success, failure, failure, success,

the three failures have ranks 1, 3, and 4, and the sum of their ranks is 8. Let W denote the sum of the ranks of x failures in n trials. If x and $n - x$ are both large and if the probability of a failure is the same for the entire sequence, W is approximately normal with mean $\mu_W = x(n+1)/2$ and variance $\sigma^2_W = x(n-x)(n+1)/12$. If $Z = (W - \mu_W)/\sigma_W$ is in either tail of the distribution, the null hypothesis should be rejected. If x or $n - x$ is small, statistics books give tables, or statistical computer packages calculate the exact tail probability.

The data of Example 6.10 show 12 failures in 63 demands. The first failure was on the first demand (01/05/87), so that failure has rank 1. The next was on the sixth demand, so that failure has rank 6. Two demands occurred on 03/19/90, the 36th and 37th demands. One of the two demands resulted in failure, so that failure was assigned rank 36.5, as is usual in case of ties. The sum of the ranks of the failures is 321.5, and Z can be calculated to equal −1.09. This is the 13.8th percentile of the normal distribution. Because Z is not in either tail, H_0 is not rejected.

6.3.3.3 Independence of Outcomes

The second assumption for binomial data is that the outcomes of different demands be independent — a success or failure on one demand does not influence the probability of failure on a subsequent demand.

Outcomes can be dependent in many ways, and some of them must be addressed by careful thinking rather than by statistical data analysis. The analyst or the study

team should consider possible common-cause mechanisms, and examine the data to see if many common-cause failures occurred. If common-cause failures form a noticeable fraction of all the failures, the analyst should probably divide the independent failures and the common-cause failures into separate data sets, and separately estimate the probabilities of each kind of failure.

The rest of this section is less important on the first reading than other sections. Some readers may wish to skip directly to Section 6.3.3.4.

If demands occur in sequence, it is natural to consider serial dependence, in which the occurrence of a failure on one demand influences the probability of a failure on the next demand. Some people believe that hits in baseball occur this way, that a slump or streak can persist because of a batter's attitude, which is influenced by how successful he has been recently. In the context of hardware failures, suppose that failures are sometimes diagnosed incorrectly, and therefore repaired incorrectly. Immediately after any failure, the probability of failure on the next demand is higher, because the first failure cause may not have been truly corrected. In such a case, the failures would tend to cluster, rather than being uniformly scattered among the successes. A cumulative plot, such as that in Figure 6.39, can be inspected for such clusters.

If the question of independence is restricted to successive outcomes — outcome i ! 1 versus outcome i — the data can be analyzed by a 2×2 contingency table. Let y_i be the outcome on demand i, either success or failure. Let x_i be the outcome on demand i ! 1. The possible values of successive outcomes (x_i, y_i) are (S, S), (S, F), (F, S), and (F, F).

To put this in more familiar language, let p denote the probability of a failure, and consider two kinds of demands, those when the previous outcome (x) was a failure and those when the previous outcome was a success. The null hypothesis is

H_0: p is the same on both kinds of demands .

Perform the usual chi-squared test of H_0 based on a contingency table.

Example 6.10 results in the contingency table shown in Table 6.15. Although the chi-squared approximation should be acceptable, it is preferable to use Fisher's exact test for a 2×2 table. The p-value reported by SAS for Fisher's exact test is 0.67. This large p-value shows that the data are very consistent with the hypothesis of independence of successive outcomes. Because the data come from the entire industry, independence is entirely reasonable.

Table 6.15 Contingency table for successive outcomes in Example 6.10.

	x = F	x = S	Total
y = F	1	10	11
y = S	11	40	51
Total	12	50	62

6.3.3.4 Consistency of Data and Prior

If the prior distribution has mean $E_{\text{prior}}(p)$, but the observed data show x/n very different from the prior mean, the analyst must ask if the data and the prior are inconsistent, or if the prior distribution was misinformed. The investigation is similar to that in Section 6.2.3.5.

Suppose first that x/n is in the right tail of the prior distribution. The relevant quantity is the prior probability of observing x or fewer events. This is

$$\Pr(X \geq x) = \int \Pr(X \geq x|p)\, f_{prior}(p)dp \qquad (6.17)$$

where

$$\Pr(X \geq x|p) = \sum_{k=x}^{n} \binom{n}{x} p^k (1-p)^{n-x} \ . \qquad (6.18)$$

If the prior distribution is beta(", $), it can be shown that Equation 6.17 equals

$$\Pr(X \geq x) = \sum_{k=x}^{n} \binom{n}{k} \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)} \frac{\Gamma(\beta+n-k)}{\Gamma(\beta)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)}$$

where '($s) is the gamma function, a generalization of the factorial function as described in Appendix A.7.6. The name of this distribution is beta-binomial. This probability can be evaluated with the aid of software. If the prior probability is any distribution other than a beta distribution, Equation 6.17 does not have a direct analytical expression.

Just as in Sec. 6.2.3.5, one method of approximating the integral in Equation 6.17 is by Monte Carlo sampling. Generate a large number of values of p from the prior distribution. For each value of p, let y be the value of Equation 6.18, which can be calculated directly. The average of the y values is an approximation of the integral in Equation 6.17. Another method of approximating the Equation 6.17 is by numerical integration.

If the probability given by Equation 6.17 is small, the observed data are not consistent with the prior belief — the prior belief mistakenly expected p to be smaller than it apparently is.

Similarly, if x/n is in the left tail of the prior distribution of the prior distribution, the relevant quantity is the prior $\Pr(X \# x)$. It is the analogue of Equation 6.17 with the limits of the summation in Equation 6.18 going from 0 to x. If that probability is small, the prior distribution mistakenly expected p to be larger than it apparently is.

Again consider Example 6.7, one AFW failure to start in eight demands, and consider the industry prior, beta(4.2, 153.1). One easy approach is Monte Carlo simulation. Therefore, values of p were generated from the beta distribution, using the technique mentioned at the end of Section 6.3.2.5.3. That is, y_1 was generated from a gamma(4.2, 1) distribution, y_2 was generated from a gamma(153.1, 1) distribution, and p was set to $y_1/(y_1 + y_2)$.

The industry-prior mean of p is 0.027, Because the observed number of failures, one, is larger than the prior expected number, $8\times0.027 = 0.21$, we ask whether such a large failure count is consistent with the prior. The probability in question is $\Pr(X \$ 1)$. For each randomly generated p, $\Pr(X \$ 1 \mid p)$ was found, equal to $1 ! \Pr(X = 0 \mid p) = 1 ! (1 ! p)^8$. The average of these probabilities, calculated for 100,000 random values of p, was 0.192, with a standard error of 0.0003. This means that the true probability is 0.192, with negligible random error. Because this probability is not small, the data appear consistent with the prior distribution.

6.4 Failure to Change State: Standby Failure

As explained in Sec. 2.3.3, this type of failure is modeled as a failure condition that occurs at an unknown time between the most recent previous inspection, test, or demand and the present one.

Each demand corresponds to a standby time. The only thing that can be observed is whether the system is failed or not at the end of the standby period. From Equation 2.3, the probability that the system is failed at time t is

$$p = 1 ! e^{/\theta t} . \tag{6.19}$$

Suppose that x failures are observed on n demands. For any one of the failures, denote the corresponding standby time by t_i, $i = 1, ..., x$. For any one of the successes, denote the corresponding standby time by s_j, $j = 1, ..., n ! x$. All these numbers are observable in principle. Therefore, the likelihood is proportional to

$$\prod_{j=1}^{n-x} e^{-\lambda s_j} \prod_{i=1}^{x} \left(1-e^{-\lambda t_i}\right). \tag{6.20}$$

This likelihood will be treated in three distinct ways below. First, a simple special case will be considered. Second, an approximation of the likelihood will be developed and used. Finally, a way to use the exact likelihood in Bayesian analysis will be given.

First, consider a simple special case, when all the standby times are equal, say, to some number t. This can happen if all the demands are test demands at equally spaced intervals. In this case, the probability of failure on demand is the same for each demand, the quantity p given by Equation 6.19. Therefore, the number of failures in n demands is binomial(n, p). The analysis methods of Section 6.3 can all be used – Bayesian or frequentist estimation of p and all the methods of model validation. At the very end of the analysis, the conclusions in terms of p should be translated into conclusions in terms of θ, by solving Equation 6.19 for

$$\theta = ! \ln(1 ! p)/t .$$

This equation for θ can be approximated as

$$\theta \; . \; p/t$$

if p is small (say, < 0.1).

This last equation shows that the MLE of θ is approximated by

$$\hat{p} / t = x / nt .$$

Here x is the number of failures and nt is the total standby time. This total standby time is approximately the total calendar time, so a simple estimate of θ is the number of failures divided by the total calendar time.

The above simple approach assumes that all the standby times are equal. If the standby times are *approximately* equal, or *nearly all* equal, it is very appealing to use the above technique, calling it an adequate approximation. If, instead, the standby times differ greatly, one of the two approaches given below can be used. The first uses an approximation of the likelihood, and the second is an exact Bayesian method.

An approximation of the exact likelihood given in Equation 6.20 can be developed as follows. It is well known that

$$1 - \exp(-\lambda t_i) \approx \lambda t_i \ .$$

This is the first order Taylor-series approximation, and is valid when λt_i is small. The error is on the order of $(\lambda t_i)^2$. A second-order approximation is less well known, but it is not hard to show that

$$1 - \exp(-\lambda t_i) \approx \lambda t_i \exp(-\lambda t_i/2) \ .$$

That is, the two quantities on the left and right of the $\approx$ have the same second order Taylor expansions, and they differ only by a term of order $(\lambda t_i)^3$. Therefore, the likelihood in Equation 6.20 is approximately equal to

$$\exp\left(-\sum_{j=1}^{n-x} \lambda s_j\right)\left(\prod_{i=1}^{x} \lambda t_i\right) \exp\left(-\sum_{i=1}^{x} \lambda t_i / 2\right) \ .$$

This is proportional to

$$e^{-\lambda t} \lambda^x$$

where

$$t = \sum_{j=1}^{n-x} s_j + \left(\sum_{i=1}^{x} t_i \Big/ 2\right) \ .$$

Compare this approximation of the likelihood with Equation 6.1, and see that the approximate likelihood here is proportional to the likelihood of x Poisson events in time t, where t equals the total standby time for the successes plus half the standby time for the failures.

Therefore, all the likelihood-based methods for Poisson data are approximately valid, treating the data as showing x failures in time t. The likelihood-based methods consist of maximum-likelihood estimation and all the Bayesian techniques.

The graphical methods for model validation from Section 6.2 remain valid, because they give qualitative indications and do not require a rigorous justification. The above argument also suggests that the chi-squared test of poolability in Section 6.2 can be used with the present data, because the chi-squared test is only an approximation in any case. However, no simulations to confirm this have been carried out for this handbook.

Finally, we give a different approach, an exact Bayesian method that can be used if the standby times have been recorded, based on Equation 6.20. Figure 6.40 gives a portion of a script for analyzing this type of data with BUGS, based on the exact likelihood. (See Figures 6.16 and 6.36 for similar scripts in other situations.)

```
model
{ for (i in 1:n) {
    p[i] <- 1 - exp(-lambda*t[i])
    x[i] ~ dbern(p[i])
  }
  lambda ~ dgamma(0.5, 0.00001)
}
```

Figure 6.40 Script for analyzing standby failure data exactly.

In this script, p_i is defined as $1 - \exp(-\lambda t_i)$. The random variable X_i is assigned a **Bernoulli**(p_i) distribution. This means that X_i equals 1 with probability p_i and equals 0 with probability $1 - p_i$. It is the same as a binomial distribution with $n = 1$. Finally, λ is assigned a prior distribution. In Figure 6.40, the prior distribution is chosen to be close to the Jeffreys noninformative prior for Poisson data, but any proper prior distribution could be used. BUGS requires a proper distribution, so the second parameter of the gamma distribution cannot be exactly zero. An additional required portion of the script, giving the data, is not shown in Figure 6.40.

6.5 Failures to Run during Mission

6.5.1 Estimates and Tests

This type of data can be analyzed using almost exactly the same tools as for event rates in Section 6.2. Certain tools carry over exactly, and others are approximately correct.

6.5.1.1 Likelihood-Based Methods: MLEs and Bayesian Methods

Suppose that n systems are run for their missions. (Equivalently, we might assume that a system is run for n missions.) Suppose that x of the runs result in failure, at times t_1, ..., t_x. The remaining $n - x$ runs are completed successfully, and the systems are turned off at times s_1, ..., s_{n-x}. Observe the notation: t for a failure time and s for a completed mission time. The likelihood is the product of the densities of times to failure, for the systems that fail, times the probability of no failure, for the systems that did not fail:

$$\prod_{i=1}^{x} f(t_i) \prod_{j=1}^{n-x} \Pr(\text{no failure by } s_j).$$

(As elsewhere, the capital pi denotes a product, analogous to a capital sigma for a sum.) Under the model introduced in Section 2.4, the failure rate is assumed to be constant, θ, the same for all the systems. Therefore, the time to failure has an exponential distribution. As stated in Appendix A.7.4, the density of an exponential(θ) distribution is

$$f(t) = \theta e^{-\theta t}$$

and the cumulative distribution function (c.d.f.) is

$$F(t) = 1 - e^{-\theta t}.$$

In particular, the probability of no failure by time s is $1 - F(s)$. Substitution of these values into the general expression for the likelihood results in

$$\prod_{i=1}^{x} \left[\lambda \exp(-\lambda t_i)\right] \prod_{j=1}^{n-x} \exp(-\lambda s_j)$$

$$= \lambda^x \exp\left[-\lambda\left(\sum_{i=1}^{x} t_i + \sum_{j=1}^{n-x} s_j\right)\right]$$

$$= \lambda^x \exp(-\lambda t)$$

where t is defined as $\Sigma t_i + \Sigma s_j$, the total running time.

Except for a normalizer that does not depend on θ, this is the Poisson probability of x failures in time t,

$$\exp(-\theta t)\theta^x t^x / x! .$$

Recall that Section 6.2 dealt with x failures in time t. Therefore, any statistical analysis that requires only a multiple of the likelihood is the same in Section 6.2 and here. In particular, the maximum likelihood estimate of θ is x/t. The gamma distributions form the family of conjugate priors, and any Bayesian analysis is carried out the same way for the data here, and the data in Section 6.2.

The subtle difference is that Σt_i is randomly generated here, so t is randomly generated (although if most of the systems do not fail during their missions, the random portion of t is relatively small). Also, the likelihood here is not a probability, but a combination of densities and probabilities, explaining the missing normalizer in the likelihood. These differences between this section and Section 6.2 result in small differences in the confidence intervals and the tests for poolability.

6.5.1.2 Confidence Intervals

Engelhardt (1995) recommends the following method when all the mission times equal the same value, s. The probability of a system failure before time s is

$$p = F(s) = 1 - \exp(-\theta s) . \qquad (6.21)$$

Based on x failures in n trials, find a confidence interval for p, using the methods of Sec. 6.3. Translate this into a confidence interval for θ, using Equation 6.21

$$\theta_{\text{conf, 0.05}} = -\ln(1 - p_{\text{conf, 0.05}})/s$$
$$\theta_{\text{conf, 0.95}} = -\ln(1 - p_{\text{conf, 0.95}})/s .$$

This method does not use all of the information in the data, because it ignores the times of any failures, using only the fact that there was a failure at some time before the mission time s. However, if failures are few, the loss of information is small.

Similarly, to perform tests when all the mission times are the same, for example to test whether two data subsets can be pooled, one can work with p, defined by Equation 6.19, and use the tests given in Section 6.3. The translation to θ needs to be made only at the very end of the analysis.

When the mission times are not all equal, no exact confidence interval method exists. However, Bayesian intervals can still be found, and are suggested.

6.5.1.3 Jeffreys Noninformative Prior

The Jeffreys prior can be worked out exactly, following the process given in Appendix B.5.3.1. If $\theta \times$(typical mission time) is small (say, < 0.1), then the Jeffreys

prior is approximately the same as in Section 6.2, an improper distribution proportional to $\beta^{1/2}$.

6.5.1.4 Tests for Poolability

The above arguments suggest that it is adequate to ignore the random element of t, and use the methods of Sec. 6.2, when estimating β. For testing whether subsets of the data can be pooled, the same arguments suggest that the chi-squared test of Sec. 6.2 can be used. The chi-squared distribution is only an asymptotic approximation in any case, and can probably be used even when t has a small amount of randomness, although no simulations to confirm this have been carried out for this handbook.

The rest of this section considers a diagnostic plot that was not introduced earlier.

6.5.2 Hazard Function Plot

One plot that is especially useful for failures to run is the hazard function plot. It is used to investigate whether β is constant during the entire mission. As explained in Appendix A.4.4 for a nonrepairable system, $\beta(t)\,\Delta t$ is the approximate probability that the system will fail during a time interval of length Δt, given that it has not yet failed. The precise name for β is the **hazard rate**, or **hazard function**, although it is often also called the failure rate.

Suppose that the system must run for some mission time, and the data value for that mission is either the mission time, if the system runs to the end without failing, or the failure time, if the system fails during the mission. The outcome, failure or success, is also recorded. The total data set consists of the data from a number of missions.

Now consider the possibility that β is not constant. Therefore, we write it as $\beta(t)$. An estimate of $\beta(t)\,\Delta t$ at some time t is the number of systems that failed during the interval $(t, t + \Delta t)$ divided by the number of systems that had not yet failed by time t. This leads to the following rather unsatisfactory estimate of $\beta(t)$. Divide the mission time into small intervals, each of length Δt, with the intervals so short that hardly any of them contain more than one failure time. In an interval with no recorded failures, estimate $\beta(t)$ by 0. In an interval $(t, t + \Delta t)$ with one failure, estimate $\beta(t)\,\Delta t$ by $1/n_t$, where n_t is the number of systems that had not yet failed by time t. Therefore, the estimate of $\beta(t)$ there is $1/(n_t\,\Delta t)$. For intervals with more than one failure, set the numerator to the number of failures.

This estimate consists of a number of spikes, at times when failures were observed. Because it is so unsmooth, this estimate is not at all attractive. However, it motivates a very simple estimate of the **cumulative hazard function**, defined as

$$\Lambda(t) = \int_0^t \lambda(u)\,du \ .$$

In this definition, the argument t of Λ is the upper limit of integration. Here Λ and β are related in the same way that a c.d.f. and a density are related. In particular, $\beta(t)$ is the derivative of $\Lambda(t)$.

A natural and simple estimate of $\Lambda(t)$ is a step function, which is flat except at times when failures occurred. At a time t when a failure occurred, the estimate of Λ jumps by $1/n_t$, where n_t is defined, just as above, as the number of systems that had not yet failed by time t. If exactly simultaneous failures occur, for example because of roundoff in reporting the failure times, the estimate of Λ jumps by the number of failures divided by n_t. This plot is due to Nelson (1982). The full name of the plot is the **cumulative hazard function plot**. This technique is illustrated with the following example.

Example 6.11 EDG failure-to-run times.

Grant et al. (1999b) state that 23 failures to run occurred during the EDG tests performed approximately once every 18 months. All these failures were reported by plants subject to Regulatory Guide RG1.108, and there were approximately 665 such tests performed at these plants during the study period. These tests require the EDG to run for 24 hours. Of the 23 failure reports, 19 reported the times to failure. The 19 reported times are given below, in hours.

0.17	0.33	2.67	6.00	11.50
0.23	0.35	3.00	8.00	13.00
0.25	0.93	4.00	10.00	17.78
0.33	1.18	5.50	10.00	

Grant et al. (1999b) assume that the lack of a reported time is statistically independent of the time at failure, so that the 19 reported times are representative of all 23 times.

There were approximately 665 such tests. Therefore, the cumulative hazard plot jumps by 1/665 at time 0.17 hours, by 1/664 at time 0.23 hours, and so forth, until it jumps by 1/647 at time 17.78. It is important that the duration of all the tests is known to

be 24 hours. This fact guarantees that none of the EDGs drop out early, so that after 18 failures 647 EDGs are still running. Actually, this is only approximate, because it ignores the four failures with unreported times.

The jumps are almost the same height, because 1/665 equals 1/647 to two significant digits. Therefore Grant et al. (1999b) plot the cumulative number of failures (a jump of 1 at each failure), instead of the estimated cumulative hazard function. The two graphs make the same visual impression, and the cumulative failure plot was easier to explain in the report. This plot is shown here, as Figure 6.41.

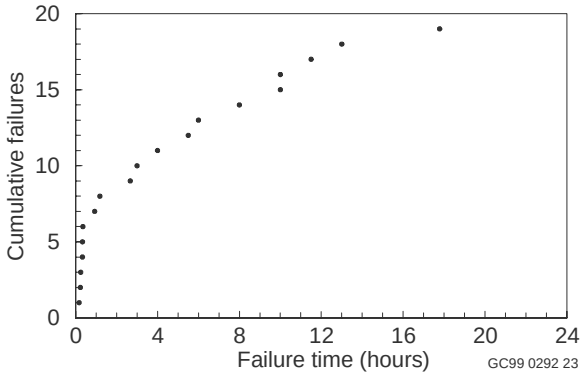


Figure 6.41 Plot of cumulative failure count, a close approximation of plot of cumulative hazard function when only a small fraction of the systems fail.

The cumulative hazard plot would differ only in that the vertical scale would be different, and the jumps would not be exactly the same size, though the jumps would be almost the same size in this example.

As explained in introductory calculus courses, when a function is graphed as a curve, the derivative of the function is the slope of the curve. Therefore, the slope of a cumulative hazard plot near time t estimates the derivative of Γ at time t. But the derivative of $\Gamma(t)$ is $\theta(t)$. Therefore, a constant slope indicates constant $\theta(t)$, and a changing slope indicates changing $\theta(t)$.

Grant et al. (1999b) note that for times less than about one half hour the slope is approximately constant, and steep. It is again constant, but less steep, from about 1/2 hour until about 14 hours, and it is smaller yet after 14 hours. Therefore, Grant et al. (1999b) estimate three values for θ, corresponding to these three time periods. They comment that the early, middle, and late failures seem to correspond in part to different failure mechanisms.

6.6 Recovery Times and Other Random Duration Times

The previous analyses have all involved a single parameter, θ or p. The analysis of duration times is different because now a distribution must be estimated, not just a single parameter.

A distribution can be estimated in many ways. If the form of the distribution is assumed, such as exponential or lognormal, it is enough to estimate one or two parameters; the parameter or parameters determine the distribution. If the form of the distribution is not assumed, the distribution can be estimated nonparametrically, or characteristics of the distribution, such as moments or percentiles, can be estimated.

To test whether data sets can be combined (pooled), both parametric tests and nonparametric tests exist. The parametric tests typically test whether the means or variances of two distributions are equal, when the distributions have an assumed form. The most common nonparametric tests test equality of the distributions against the alternative that one distribution is shifted sideways from the other.

This section is long, because so many distribution models can be assumed and because the model assumptions can be violated in so many ways. A brief outline of the section is as follows:

6.6.1 Characterization of a single distribution
- Estimation of moments, percentiles, c.d.f.s
- Fitting of four parametric models (frequentist and Bayesian parameter estimates)

6.6.2 Model validation (graphs and hypothesis tests)
- Poolability, trend
- Goodness of fit to assumed parametric models
- Consistency of data with prior for Bayesian parameter estimates

6.6.3 Nonparametric density estimation

Many of the methods will be illustrated using the data of Example 6.12, taken from Atwood et al. (1998).

This example shows the times when power could have been recovered, for plant-centered LOSP events, that is, for events not caused by grid problems or by widespread severe weather. (Real life is complicated: sometimes a plant does not restore power as quickly as it could, and the event report states when power was actually restored, and usually also when it could have been restored. The times given by Atwood et al. (1998) as "recovery times" show when power could have been restored,

if that time was reported and different from the actual recovery time.) Discussions of this example will use the terms **recovery time** and **duration** interchangeably. Momentary events (duration less than two minutes) and events with no reported duration have been excluded. For common-cause events that affected multiple units at a site, the average recovery time is used.

Example 6.12 LOSP recovery times.

Atwood et al. (1998) report 115 times of recovery of lost offsite power. The data are categorized into three possible values for plant status: T, S, and P, with meanings explained in the table below. The durations in minutes and the dates (MM/DD/YY) are shown.

P: Plant remained at power throughout LOSP event (8 times)		
6 03/01/80	113 01/18/96	385 04/11/94
45 07/25/85	147 06/03/80	1138 01/03/89
65 07/16/88	355 11/12/90	
S: Plant was shut down before and during LOSP event (62 times)		
2 06/04/84	14 11/16/84	60 06/22/91
2 08/17/87	14 02/01/81	60 06/16/89
2 06/29/89	15 04/27/81	62 07/15/80
2 05/21/94	15 12/19/84	67 03/13/91
3 06/26/93	15 10/12/93	73 08/28/85
3 10/22/84	17 04/26/83	77 03/29/92
3.5 11/21/85	17 10/14/87	97 01/08/84
4 04/22/80	20 03/23/92	120 06/05/84
4 04/04/87	22 08/24/84	120 01/16/81
4 10/20/91	24 07/29/88	127 01/20/96
5 05/03/84	24 07/29/88	132 02/27/95
8 06/24/88	29 03/20/91	136 04/08/93
9 12/26/88	29 09/16/87	140 03/20/90
10 08/01/84	29 05/14/89	155 03/05/87
10 04/28/92	35 04/02/92	163 10/08/83
10 12/23/81	37 03/21/87	240 11/14/83
11 10/04/83	37 05/19/93	240 03/07/91
11 07/24/91	37 07/09/90	335 04/29/85
12 06/22/93	43 05/07/85	917 10/21/95
12 07/19/86	53 09/11/87	1675 11/18/94
14 02/26/90	59 10/16/87	
T: Plant tripped because of LOSP event (45 times)		
2 02/28/84	20 08/21/84	90 02/12/84
4 11/21/85	20 07/16/84	90 03/29/89
4 11/17/87	20 06/27/91	90 06/17/89
5 08/16/85	24 06/15/91	95 12/31/92
6 05/03/92	25 10/03/85	95 12/31/92
10 09/10/93	29 06/22/82	95 10/16/88
10 10/12/93	38 07/17/88	96 12/27/93
11 07/26/84	40 02/11/91	100 01/28/86
13 10/07/85	45 01/16/90	106 06/03/80
14 08/13/88	45 03/25/89	118 07/23/87
15 02/16/84	46 01/01/86	118 07/23/87
15 09/14/93	57 10/19/92	277 04/23/91
19 10/25/88	60 03/21/91	330 02/06/96
20 12/12/85	60 10/22/85	388 07/14/87
20 03/27/92	62 07/15/80	454 08/22/92

The group P exists because some plants are permitted to remain at power during certain LOSP events.

Throughout this section, the random variable is denoted by T, because typically the random quantity is a duration time, such as time to recovery of the system. Several examples were given in Section 2.5.1: time until restoration of offsite power, duration of a repair time, and others. Let F denote the c.d.f. of T, $F(t) = \Pr(T \# t)$. It is assumed that n times will be observed, $T_1, T_2, \ldots, T_n$. The assumptions of Section 2.5.2 are repeated here.

- The T_i s are independent,
- Each T_i has the c.d.f. $F(t)$.

A data set satisfying these assumptions is called a **random sample** from the distribution. Sometimes the T_i s are called **independent identically distributed** (i.i.d.). The term random sample can refer either to the random variables (T_is) or to the observed values, $t_1, t_2, \ldots, t_n$. The data are used to estimate properties of the distribution. This can also be described as estimating properties of the population, where the **population** is the infinite set of values that could be randomly generated from the distribution.

6.6.1 Characterization of Distribution

6.6.1.1 Nonparametric Description

The tools in this subsection are called nonparametric because they do not require any assumption about the form of the distribution. For example, the distribution is not assumed to be lognormal, exponential, or any other particular form.

6.6.1.1.1 Moments

To estimate the population mean μ or a population variance σ^2, two simple estimators are the **sample mean**, defined as

$$\bar{T} = \frac{1}{n}\sum_{i=1}^{n} T_i$$

and the **sample variance**, defined as

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (T_i - \bar{T})^2 \quad .$$

The sample mean and sample variance are known to be **unbiased** for the population mean and variance, respec-

tively. In other words, $E(\bar{T}) = \mu$ and $E(S^2) = \sigma^2$. These statements are true regardless of the distribution F, requiring only the assumptions of a random sample. The **sample standard deviation**, S, is the square root of the sample variance. When defining S^2 some authors use n in the denominator instead of $n - 1$, with corresponding adjustment of formulas that involve S, but this handbook uses the above definition consistently, both here and in Appendix B. In applications with computer packages, note which definition is used and make any necessary adjustments to formulas in this handbook.

These are all-purpose estimators, but they are not the only possible estimators. For example, the variance of an exponential distribution is the square of the mean. Therefore, a good estimator of the variance would be the square of the estimator of the mean. This estimator relies heavily on the assumption of exponentiality, whereas the above estimators make no such assumptions. General principles of estimation are discussed in Appendix B.4.1.

6.6.1.1.2 Percentiles

To estimate percentiles of a distribution, it is useful to put the data in ascending order from the smallest to the largest observation. The recovery times in Example 6.12 have been arranged this way. The variables obtained by ordering the random sample are called the **order statistics**, and are denoted by $T_{(1)} \le T_{(2)} \le \cdots \le T_{(n)}$. The observed values are written $t_{(1)} \le t_{(2)} \le \cdots \le t_{(n)}$. Some important estimates based on the order statistics are the sample median, other sample percentiles, and the sample range. The general definition of the $100q$th sample percentile, where $0 < q < 1$, is a number t such that the fraction of observations that are $\le t$ is at least q and the fraction of observations that are $\ge t$ is at least $1 - q$.

For example, the **sample median** is defined to be t such that at least half (because $q = 0.5$) of the observations are $\le t$ and at least half (because $1 - q = 0.5$) are $\ge t$. This boils down to the following. If n is odd, the sample median is the "middle" order statistic, $t_{(m)}$ with $m = (n + 1)/2$. If n is even, with $m = n/2$, there is no unique "middle" order statistic. Any number between the two middle order statistics, $t_{(m)} \le t \le t_{(m+1)}$, could be used. However, nearly everyone uses the average of the two middle order statistics $(t_{(m)} + t_{(m+1)})/2$ as "the" sample median.

The other sample percentiles are defined similarly, with some averaging of two order statistics if necessary. Note that the sample 90th percentile is $t_{(n)}$ if $n < 10$, the sample 95th percentile is $t_{(n)}$ if $n < 20$, and so forth.

Order statistics that are sometimes used are: the lower and upper **quartile**, defined as the 25th and 75th percentiles; percentiles that include most of the distribution, such as the 5th and 95th percentiles; and the extremes, $t_{(1)}$ and $t_{(n)}$. The **interquartile range** is the upper quartile minus the lower quartile. The **sample range** is the difference between the largest and smallest ordered observations, $t_{(n)} - t_{(1)}$. Be careful with interpretation. As data continue to be collected, the sample interquartile range stabilizes at the interquartile range of the distribution, but the sample range does not stabilize at all — it just grows every time a new t is observed that is outside the former observations.

The sample median has the advantage of not being strongly influenced by extreme observations. The sample mean, on the other hand, can be strongly influenced by even one extreme data value. The sample variance is even more sensitive to extreme values, because it is based on squared terms. Therefore, the sample standard deviation, defined as the square root of the sample variance, is also sensitive to extreme terms. Other measures of dispersion, such as the interquartile range, are much less sensitive to extreme values. In general, sample percentiles are much less sensitive to extreme observations than are sample moments.

The recovery times of Example 6.12 have sample moments and percentiles given in Table 6.16.

Table 6.16 Statistics based on the recovery times (minutes) of Example 6.12.

	P	S	T
n	8	62	45
Stand. deviation	373.2	241.4	99.9
95th %ile	1138	240	330
75th %ile (upper quartile)	370	73	95
Mean	281.75	92.3	73.4
50th %ile (median)	130	24	40
25th %ile (lower quartile)	55	10	15
5th %ile	6	2	4

For the P group, the sample median is taken as the average of the two middle numbers. Even though

the S group has an even number of observations, its sample median is unique, because $t_{(31)}$ and $t_{(32)}$ happen to be equal. The T group has an odd number of observations, so its sample median is unique, $t_{(23)}$.

The S group has one very extreme value, which influences the moments. The sample mean for this group is larger than the upper quartile — someone who considers the mean to be "the" average could say that more than 75% of the observed times are below average. Such a thing can happen with a skewed distribution. This is one reason why many people prefer percentiles to moments for describing a skewed distribution.

There are situations in which some of the times are not observed. Section 6.5 dealt with such a situation, when the times of interest were times of EDG failure to run, and not all these times were reported. In the present section, nearly all the times are assumed to be observed, with no systematic bias in which times fail to be observable.

6.6.1.1.3 The Empirical Distribution Function

An estimate of $F(t)$ called the **empirical distribution function** (EDF) is defined as follows: For an arbitrary value of $t > 0$, define

$\hat{F}(t)$ = (Number of observations # t) / n.

The EDF is a step function. It increases by $1/n$ at each observed time if all observations are distinct. More generally, if there are m times equal to t, $\hat{F}(t)$ has a positive jump of m/n at t.

In some settings the function

$1 - F(t) = \Pr(T > t)$

is of interest. If T is the time until failure, $1 - F(t)$ is called the **reliability function**, $R(t)$, in engineering contexts, and the **survival function**, $S(t)$, in medical contexts. A suitable word remains to be coined when T is the time until recovery or repair. The empirical reliability function, or the empirical survival function, is defined as 1 minus the EDF. Anything that can be done with F can be translated in terms of $1 - F$, so this discussion will only consider F.

With a little mental exercise, the EDF can be expressed in familiar terms. For any particular t, let p denote $F(t) = \Pr(T \# t)$. In the data, define a "demand" to be the generation of an observed time, and define the ith observation t_i to be a "failure" if $t_i \# t$. By the assumptions for a random sample, any observation has probability p of being a failure, and the outcomes (failures or successes) are statistically independent of each other. By its definition, $\hat{F}(t)$ is the number of failures divided by the number of demands, which is $\hat{p}$, as indicated in Section 6.3.1. Therefore, $\hat{F}(t)$ is an unbiased estimator of $F(t)$ at any t. It is close to $F(t)$ when the number of observations is large, and a confidence interval for $F(t)$ can be constructed, the familiar confidence interval for p.

Figure 6.42 shows the EDF based on the data in Example 6.12 for group T.

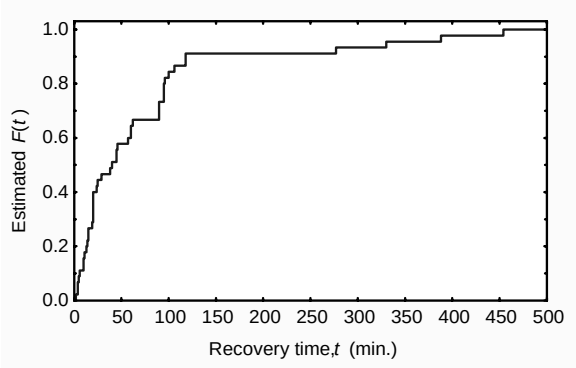


Figure 6.42 Empirical distribution function (EDF) for the data from group T in Example 6.12.

6.6.1.1.4 Histogram Estimate of the Density

The eye smooths the EDF, compensating for its jagged form. To accomplish the same sort of smoothing for a density estimate, group the observed times into bins of equal width, count the number of observations in each bin, and plot the histogram, a form of bar chart with the height of each bin equal to the number of observations in the bin. The histogram is *proportional* to an estimate of the density. Some software packages can rescale the height of the histogram so that the total area equals 1, making it a true density estimate. Books and Ph. D. theses have been written on density estimation, and some modern density estimators are quite sophisticated. A few such are given in Section 6.6.3. Nevertheless, the lowly histogram is often adequate for PRA purposes.

Figures 6.43 and 6.44 show two histograms for the data from the above EDF, using two different bin widths. The analyst must decide what bin width gives the most reasonable results, based on belief about how smooth or ragged the true density might

be. Most people would judge Figure 6.44 to be too rough, and would therefore choose wider bins.

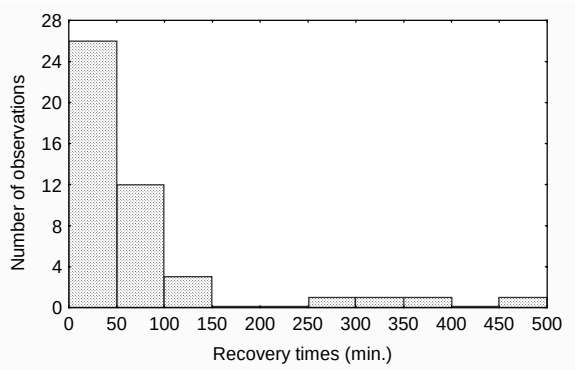


Figure 6.43 Histogram of the data from group T in Table 6.16, with bin width 50.

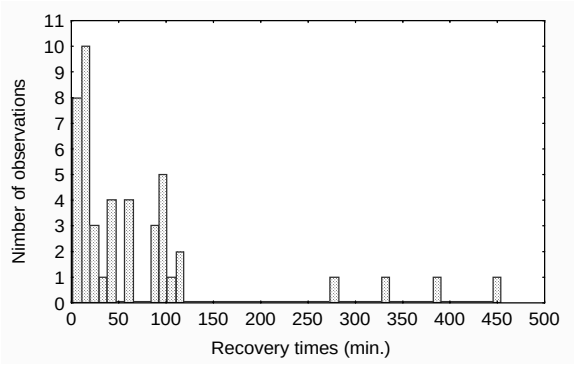


Figure 6.44 Histogram of same data, with bin width 10.

6.6.1.2 Fitting a Parametric Distribution

Sometimes it is desirable to fit some assumed distributional form to data. This subsection gives estimators if the assumed distribution is lognormal, exponential, gamma, or Weibull. Bayesian and non-Bayesian estimates are given, with much of the latter taken from an INEEL report by Engelhardt (1996).

6.6.1.2.1 Lognormal Distribution

This model assumes that T has a lognormal distribution, or equivalently, that $\ln T$ has a normal(μ, σ^2) distribution. Define $X = \ln T$.

Frequentist Estimates. The usual estimates of μ and σ^2 are:

$$\bar{x} = \frac{1}{n}\sum_i x_i$$

and

$$s_X^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2 .$$

These estimates have the same form as those given in Section 6.6.1.1.1 for the mean and variance of T, but these are for $\ln T$. Calculate the estimates of μ and σ^2 to determine the estimated normal distribution of $\ln T$, which determines the estimated lognormal distribution of T. Note that the sample variance is defined with $n - 1$ in the denominator, although some authors use n in the definition and slightly different formulas below.

The material below is presented in many statistics books, based on the fact that $\ln T$ has a normal distribution. The distribution of $(n-1)S_X^2 / \sigma^2$ is chi-squared with $n - 1$ degrees of freedom. It follows that a two-sided $100(1 - \alpha)\%$ confidence interval for σ^2 is

$$\left((n-1)s_X^2 / \chi^2_{1-\alpha/2}(n-1), (n-1)s_X^2 / \chi^2_{\alpha/2}(n-1)\right).$$

Here $\chi_q^2(n-1)$ is the q quantile, that is, the $100q$ percentile, of the chi-squared distribution with $n-1$ degrees of freedom.

The distribution of $\bar{X}$ is normal(μ, σ^2/n). If σ^2 is known, it follows that a $100(1 - \alpha)\%$ confidence interval for μ is

$$\bar{x} \pm z_{1-\alpha/2}\sigma / \sqrt{n} ,$$

where $z_{1-\alpha/2}$ is the $100(1 - \alpha/2)$ percentile of the standard normal distribution. For example, $z_{0.95}$ gives a two-sided 90% confidence interval.

In the more common case that both μ and σ^2 are unknown, use the fact that

$$(\bar{X} - \mu) / (S_X / \sqrt{n})$$

has a Student's t distribution with $n-1$ degrees of freedom. It follows that a $100(1 - \alpha)\%$ confidence interval for μ is

$$\bar{x} \pm t_{1-\alpha/2}(n-1)s_X / \sqrt{n} ,$$

where $t_{1-\alpha/2}(n-1)$ is the $1 - \alpha/2$ quantile of the Student's t distribution with $n-1$ degrees of freedom. For example, $t_{0.95}(n-1)$ gives a two-sided 90% confidence interval. Percentiles of Student's t distribution are tabulated in Appendix C.

Bayesian Estimation. Bayesian estimates are given here.

Conjugate Priors. The conjugate priors and update formulas are presented by Lee (1997, Sec. 2.13). They depend on four prior parameters, denoted here as d_0, F_0^2, n_0, and μ_0. The notation here follows the notation used elsewhere in this handbook. It is not the same as Lee's. Quantities with subscripts, such as F_0^2 or d_1, are numbers. Quantities without subscripts, σ^2 and μ, have uncertainty distributions.

It is useful to think of having d_0 degrees of freedom, corresponding to $d_0 + 1$ prior observations for estimating the variance, and a prior estimate F_0^2. More precisely, let the prior distribution for $\sigma^2/(d_0F_0^2)$ be inverted chi-squared with d_0 degrees of freedom. That is, $d_0F_0^2/\sigma^2$ has a chi-squared distribution with d_0 degrees of freedom. Therefore it has meand_0, and therefore the prior mean of $1/\sigma^2$ is $1/F_0^2$. (See Appendix A.7.7 for more information on the inverted chi-squared distribution.)

An alternative notation for the above paragraph would define the **precision** $J = 1/\sigma^2$, and the prior precision $J_0 = 1/F_0^2$. Then the prior distribution of d_0J/J_0 is chi-squared with d_0 degrees of freedom. Although we shall not use this parameterization, it has adherents. In particular, BUGS (1995) uses J instead of σ^2 as the second parameter of the normal distribution; see Spiegelhalter et al. (1995).

Conditional on σ^2, let the prior distribution for μ be normal with mean μ_0 and variance σ^2/n_0. This says that the prior knowledge of μ is equivalent to n_0 observations with variance σ^2. It is not necessary for n_0 to have any relation to d_0.

The Bayes update formulas are

$d_1 = d_0 + n,$
$n_1 = n_0 + n,$
$\mu_1 = (n_0\mu_0 + n\bar{x}) / n_1$ and

$$\sigma_1^2 = \left[d_0\sigma_0^2 + (n-1)s_X^2 + \frac{n_0 n_1}{n_0 + n_1}(\mu_0 - \bar{x})^2 \right] / d_1 .$$

Here the subscript 1 identifies the posterior parameters. The posterior distributions are given as follows. First, $\sigma^2/(d_1F_1^2)$ has an inverted chi-squared distribution with d_1 degrees of freedom. That is, the posterior mean of $1/\sigma^2$ is $1/F_1^2$, and a two-sided 100(1 ! ")% credible interval for σ^2 is

$$\left(d_1\sigma_1^2 / \chi^2_{1-\alpha/2}(d_1), d_1\sigma_1^2 / \chi^2_{\alpha/2}(d_1) \right).$$

Conditional on σ^2, the posterior distribution of μ is normal(μ_1, σ^2/n_1). Therefore, conditional on σ^2, a two-sided 100(1 ! ")% credible interval for μ is

$$\mu_1 \pm z_{1-\alpha/2}\sigma / \sqrt{n_1} .$$

The marginal posterior distribution of μ, that is, the distribution that is not conditional on σ^2, is as follows. The expression

$$(\mu - \mu_1) / (\sigma_1 / \sqrt{n_1})$$

has a Student's t distribution with d_1 degrees of freedom. It follows that a 100(1 ! ")% credible interval for μ is

$$\mu_1 \pm t_{1-\alpha/2}(d_1)\sigma_1 / \sqrt{n_1} .$$

Noninformative Prior. The joint noninformative prior for (μ, σ^2) is proportional to $1/\sigma^2$. Lee (1997, Sec. 2.13) presents this prior, as do Box and Tiao (1973, Sec. 2.4). Lee points out that when d_0 = ! 1, $n_0 = 0$, and $F_0^2 = 0$, the conjugate prior distribution reduces to the noninformative prior. In the formulas just given above, $n_1 = n$, $d_1 = n$! 1, $\mu_1 = \bar{x}$, and $F_1 = s_X$. The credible intervals then agree numerically with the confidence intervals given above.

Possible Further Analyses. Some data analyses require only the posterior distribution of one or both parameters. In that case, use the above posterior distributions, with either an informative or noninformative prior. Other analyses require more, such as simulation of a set of lognormal times T or a credible interval for the mean of T. If so, simulation of the quantity of interest is a useful technique. Begin each case of the simulation by generating a value of σ^2 from its posterior distribution. Then generate a value of μ from its distribution conditional on σ^2. Then do whatever is required next to obtain the quantity of interest: generate a random value of T from the lognormal(μ, σ) distribution, or calculate $E(T) = \exp(\mu + \sigma^2/2)$, or calculate whatever else is needed. Save the quantity of interest produced in this way. Repeat this process as many times as needed to obtain a sample that accurately represents the distribution of interest.

Model Validation. Model validation is discussed in general in Section 6.6.2. Many of the methods given there are applicable to any assumed distribution. Some methods, however, have been developed just for the normal and lognormal distributions. They are contained in Sections 6.6.2.1.2, 6.6.2.2.2, and 6.6.2.3.2.

6.6.1.2.2 Exponential Distribution

The exponential distribution is related to a Poisson process, because the times between successive events in a Poisson process have an exponential distribution.

The exponential distribution is presented in Appendix A.7.4, with two possible parameterizations. The first uses $\theta = 1/E(T)$, and the second uses $\beta = 1/\theta = E(T)$. In data analysis, sometimes one parameter seems more natural and convenient than the other. In the two parameterizations, the likelihood function is

$$\theta^{n}\exp(-\theta\Sigma t_i)$$

or

$$\beta^{-n}\exp(-\Sigma t_i/\beta) .$$

As a function of θ, the likelihood function here is proportional to the likelihood function given by Equation 6.1 for Poisson data. (Replace x in Equation 6.1 by n and t by $\Sigma_i t_i$.) Therefore, many of the results below are similar to or identical to the results in Section 6.2 for Poisson data.

Frequentist Estimation. It can be shown that the MLE of β is the sample mean, $\bar{t}$. Therefore, to estimate the distribution, estimate β by $\bar{t}$. This determines the estimated exponential distribution. The corresponding estimate of $\theta = 1/\beta$ is $1/\bar{t}$.

For a $(1-\alpha)$ confidence interval, or equivalently a $100(1-\alpha)\%$ confidence interval, the lower limit for θ is

$$\theta_{\text{conf},\,\alpha/2} = \frac{\chi^2_{\alpha/2}(2n)}{2\Sigma t_i}$$

and the upper limit is

$$\theta_{\text{conf},\,1-\alpha/2} = \frac{\chi^2_{1-\alpha/2}(2n)}{2\Sigma t_i} .$$

(See Martz and Waller 1991.) Confidence limits for $\beta = 1/\theta$ are obtained by inverting the confidence limits for θ. For example, the lower confidence limit for β equals 1 divided by the upper confidence limit for θ.

Bayesian Estimation. Now consider Bayesian estimation.

Conjugate Prior. The gamma distribution is a conjugate prior for θ. That is, let $t_1, \ldots, t_n$ be independent observations from an exponential(θ) distribution. Let the prior distribution of θ be gamma(α_0, β_0). This uses the same parameterization as when θ is a Poisson parameter (Section 6.2.2.4), so that β_0 has units of time and the prior mean of θ is α_0/β_0. A direct calculation shows that the posterior distribution of θ is also gamma, with posterior parameters

$$\alpha_1 = \alpha_0 + n$$
$$\beta_1 = \beta_0 + \Sigma t_i .$$

The subscript 1 identifies the posterior parameters. The prior parameters have a simple intuitive interpretation – the prior information is "as if" α_0 duration times had been observed with total value β_0.

The percentiles of the posterior distribution are given by

$$\lambda_p = \frac{\chi^2_p(2\alpha_1)}{2\beta_1} .$$

Therefore, for example, a two-sided 90% credible interval has end points

$$\lambda_{0.05} = \frac{\chi^2_{0.05}(2\alpha_1)}{2\beta_1}$$

and

$$\lambda_{0.95} = \frac{\chi^2_{0.95}(2\alpha_1)}{2\beta_1} .$$

There are two possible ways to perform the corresponding analysis in terms of β. (a) One way is to perform the above analysis in terms of θ, and then translate the answer into answers for $\beta = 1/\theta$. Be careful when doing this. The percentiles translate directly, with the $100p$ percentile $\beta_p = 1/\theta_{1-p}$. For example, $\beta_{0.95} = 1/\theta_{0.05}$. The moments do not translate directly, however. For example, the posterior mean of β is $\beta_1/(\alpha_1 - 1)$, not 1 divided by the mean of θ. (b) The other way is to let β have an inverted gamma distribution. This distribution is defined in Appendix A.7.7.

Either analysis gives exactly the same results. The second approach is just a disguised version of the first approach, using a different distribution to avoid introduction of the symbol θ.

Noninformative Prior. The Jeffreys noninformative prior for θ can be expressed as a gamma(0, 0) distribution. This is an improper distribution, that is, it does not integrate to 1, but it does result in proper

posterior distributions as long as some data have been observed. Note, this prior is slightly different from the Jeffreys prior when the data have a Poisson distribution. When the gamma(0, 0) prior is used with exponential data, the posterior parameters reduce to

$$\alpha_{post} = n$$
$$\beta_{post} = \Sigma t_i .$$

Then the Bayes posterior credible intervals are numerically equal to the confidence intervals. If the purpose of a "noninformative" prior is to produce intervals that match confidence intervals, this purpose has been perfectly accomplished.

Discussion. The above work has illustrated some facts that are generally true. When the observations have a discrete distribution, such as Poisson or binomial, the so-called noninformative priors do not produce credible intervals that exactly match confidence intervals. This is related to the fact that confidence intervals from discrete data do not have exactly the desired confidence coefficient. Instead, they are constructed to have *at least* the desired long-run coverage probability. The situation is different when the observations are continuously distributed, as in the present case with exponentially distributed times. In this case, the confidence intervals have exactly the desired long-run coverage probability, and posterior credible intervals, based on the noninformative prior, are numerically equal to the confidence intervals.

Nonconjugate priors can also be used. The procedure is similar to that in Section 6.2.2.6, but now uses the exponential likelihood given above. Therefore, it is not discussed here.

Model Validation. Model validation is discussed in general in Section 6.6.2. Many of the methods given there are applicable to any assumed distribution. A few methods, however, have been developed just for the exponential distribution. They are mentioned in Sections 6.6.2.3.1 and 6.6.2.4.1.

6.6.1.2.3 Gamma Distribution

The distribution of T is gamma(α, β) if the density is

$$f(t) = \frac{1}{\tau^{\alpha}\Gamma(\alpha)} t^{\alpha-1} e^{-t/\tau} .$$

Note, this is a different parameterization from the previous section and from Equation 6.4. This parameterization is related to the earlier parameterization by the relation $\beta = 1/\lambda$. In the present context, t and β both have units of time.

The MLEs of the parameters are given by Bain and Engelhardt (1991, p. 298) or by Johnson et al. (1994, Sec. 17.7). They are the solutions of the equations

$$\tau = \bar{t} / \alpha$$
$$\ln(\alpha) - \psi(\alpha) = \ln(\bar{t} / \tilde{t}) , \qquad (6.23)$$

where $\psi(u) = \Gamma'(u)/\Gamma(u)$ is the digamma function, calculated by some software packages, and

$$\tilde{t} = \exp\left[(1/n)\Sigma \ln t_i\right],$$

is the geometric mean of the observed times. Equation 6.23 must be solved for α by numerical iteration. Bain and Engelhardt (1991, p. 298) give a table of approximate solutions, which may be interpolated.

The MLEs of the two parameters determine the estimated gamma distribution.

Bayes estimation is complicated because the gamma distribution, like the lognormal distribution, has two parameters, and these two parameters must have a joint distribution. Martz and Waller (1991, Sec. 9.5.2) cite Lwin and Singh (1974) for an analysis that was feasible in the 1970s. A simpler approach today would use the freely available software package BUGS (1995), described in Section 6.2.2.7, Section 8.3.3.3, and elsewhere in this handbook. BUGS is designed for models with many unknown parameters, and should make short work of a model with only two. The joint prior distribution would not need to be conjugate.

6.6.1.2.4 Weibull Distribution

A three-parameter Weibull distribution is given in Appendix A.7.5. A two-parameter form of the Weibull distribution is given here, by setting the location parameter γ to zero. The density is

$$f(t) = (\beta / \alpha)(t / \alpha)^{\beta-1} \exp\left[-(t / \alpha)^{\beta}\right] .$$

As with the gamma distribution, the maximum likelihood equations do not have closed-form solutions. The estimates must be found by iteratively solving

$$\frac{\Sigma t_i^{\beta} \ln(t_i)}{\Sigma t_i^{\beta}} - \frac{1}{\beta} - \frac{1}{n}\Sigma \ln t_i$$

and

$$\alpha = \left(\frac{1}{n} \Sigma t_i^{\beta} \right)^{1/\beta} .$$

Zacks (1992, Section 7.5) gives the following simple method for solving the first equation. Begin with $\hat{\beta}_0 = 1$. Then repeatedly solve the equation

$$\hat{\beta}_{n+1} = 1 \Big/ \left[\frac{\Sigma t_i^{\hat{\beta}_n} \ln(t_i)}{\Sigma t_i^{\hat{\beta}_n}} - \frac{1}{n} \Sigma \ln t_i \right]$$

with $n = 0, 1, 2, \ldots$ The value of $\hat{\beta}_n$ converges quickly to the MLE $\hat{\beta}$. Then set

$$\hat{\alpha} = \left(\frac{1}{n} \Sigma t_i^{\hat{\beta}} \right)^{1/\hat{\beta}} .$$

For more information, see Zacks (1992) or Bain and Engelhardt (1991).

Alternatively, a simple approximate graphical estimate is based on the hazard function. Plots of the cumulative hazard were discussed in Section 6.5.2. It can be shown that the cumulative hazard function of the Weibull distribution is

$H(t) = (t/\,'')^{\$}$.

Therefore, estimate the cumulative hazard function as explained in Section 6.5.2, by jumping at each observed time, with the size of the jump equal to 1 divided by the number of times that have not yet been equalled or exceeded. The jump at $t_{(1)}$ is $1/n$, the jump at $t_{(2)}$ is $1/(n - 1)$, and so forth until the final jump at $t_{(n)}$ is 1. Call this estimate $\hat{H}(t)$. The equation for the Weibull cumulative hazard function can be rewritten as

$$\log H(t) = \$ \log t - \$ \log\,'', \qquad (6.24)$$

which is linear in log*t*. Therefore, plot $\log[\hat{H}(t)]$ against log*t*, that is, plot $\hat{H}(t)$ against *t* on log-log paper, and fit a straight line to the plot by eye. Pick a point on the line and substitute those values of *t* and $\hat{H}(t)$ into Equation 6.24. This is one equation that $\$$ and log $''$ must satisfy. Pick a second point on the line and obtain a second equation in the same way. Solve those two equations for $\$$ and log $''$, thus obtaining estimates of $\$$ and $''$. In the calculations, it does not matter whether natural logarithms or logarithms to base 10 are used, as long as the same type is used everywhere.

This plot also gives a diagnostic test of whether the Weibull distribution is appropriate. The degree to which the plotted data follow a straight line indicates the degree to which the data follow a Weibull distribution.

Just as in Sections 6.6.1.2.1 and 6.6.1.2.3, Bayes estimation is complicated here by the multiple parameters. Martz and Waller (1991, Sec. 9.1) cite a number of early papers using various prior distributions. However, the easiest Bayesian approach today would be to assign convenient diffuse priors to the parameters and use BUGS (1995), described in Section 6.2.2.7, Section 8.3.3.3, and elsewhere in this handbook.

6.6.2 Model Validation

This section considers several topics. First, the usual investigations of the model assumptions are considered: whether subsets of the data all correspond to the same distribution, whether the distribution changes with time, and whether the times are serially correlated instead of statistically independent. In addition, the distribution may have been modeled by some parametric form, so the goodness of fit is investigated. Finally, if parameters have been estimated in a Bayesian way, the consistency of the data with the prior must be investigated.

The order described above follows the actual order of analysis. First, the analyst would check to see what data subsets can be pooled and whether the usual assumptions seem to be satisfied. Only then would it be appropriate to try to fit some standard distribution to the data.

6.6.2.1 Poolability of Data Sources

To illustrate the methods here, this subsection will consider the three groups of data in Example 6.12, corresponding to three conditions of the plant during the LOSP event. As elsewhere in this chapter, graphical methods are considered first, and statistical tests second.

6.6.2.1.1 Graphical Methods

A simple, graphical method of comparison is to overlay the EDFs for the different data subsets on a single graph. Then, look to see if the EDF are intertwined, indicating that the subsets may be pooled, or if they are

separated, shifted sideways from each other, indicating that the data subsets may not be pooled. This method is simple, but the graph can become very cluttered, especially if a moderate or large number of subsets must be compared. The same comment can be made for comparing separate histograms of the data subsets.

A graph that has come into common use is the **box-and-whisker plot**, or **box plot**. The lower and upper edges of the box are the lower and upper quartiles of the data. Thus, the box can be thought of as containing half the data, with 1/4 of the remaining data on each side. The median is marked somehow. The "whiskers" are two lines coming out of the box and going out to cover the range of most of the data, up to 1.5 times the interquartile range in each direction. A few outlying points are plotted individually.

Figure 6.45 shows a box plot of the group T data from Example 6.13 generated using the STATISTICA (1995) software. The median is marked by a small square in the box. The software documentation does not give a precise definition of the difference between an outlier and an extreme point. Also, this release of the software seems to have a small bug, in that the maximum (excluding outliers) is labeled as 11, when it should be 118.

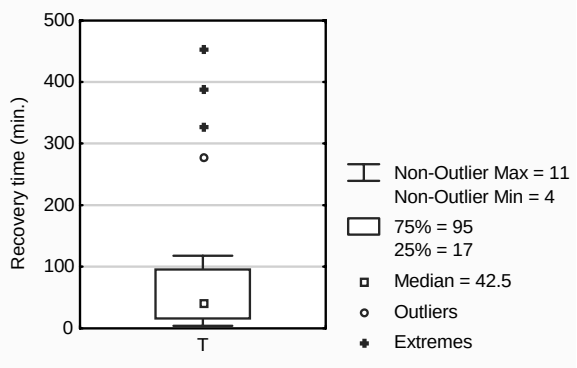


Figure 6.45 One form of a box plot. The box shows the lower and upper quartiles, with the median marked. The whiskers show most of the range, from 4 to 118, and individual outlying points are plotted.

Figure 6.46 shows the same box plot as drawn by a different software package, SAS/INSIGHT (1995). As before, the box shows the lower and upper quartiles, and the median is marked, this time with a stripe. Points beyond the whiskers are shown as individual dots.

Box plots were invented by Tukey (1977), and are still being modified according to individual taste. Any form of the plot that is produced by a convenient software package is probably adequate.

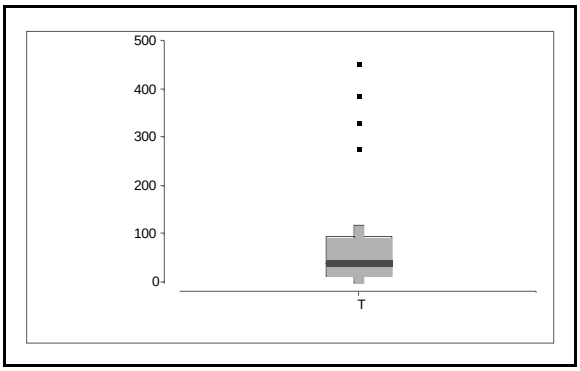


Figure 6.46 A different style box plot of the same data. The box shows the upper and lower quartiles, with the median indicated by a stripe. The whiskers show much of the range, with dots marking outliers.

The example here is typical, in that the data are skewed, and the most obvious feature of the box plots given here is the long distance from the box to the largest value. Box plots are supposed to focus on the bulk of the data, with only moderate attention given to the extremes. Therefore, there are visual advantages to transforming skewed data by taking logarithms. Therefore, all the remaining box plots shown in this section will use $\log_{10}$(recovery time) instead of the raw times.

Figure 6.47 shows side-by-side box plots of the three data subsets in Example 6.12. Incidentally, the box plot of log(time) is different from the box plot of time plotted on a logarithmic axis — the logarithms of large times tend not to be considered as outliers. This can be seen by comparing Figure 6.45 with the group-T portion of Figure 6.47.

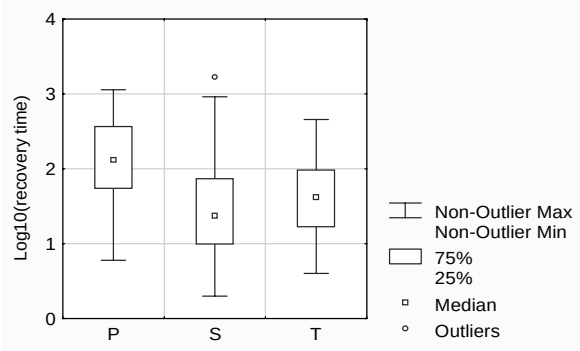


Figure 6.47 Side-by-side box plots of the three groups of data from Table 6.16, based on $\log_{10}$(recovery time).

Figure 6.47 shows that group P seems to have somewhat longer recovery times than the other groups. There seems to be little difference between groups S and T. Tests will be given below to investigate whether this visual impression is correct.

6.6.2.1.2 Statistical Tests

Tests Based on Normality. Warning: these tests are only valid if normality or lognormality can be assumed. If each data subset corresponds to a lognormal distribution, work with $X = \log(T)$. Either natural logs or base-10 logs can be used, because $\log_{10}(T) = \ln(T)/\ln(10)$, so both are normally distributed if either is.

When X has a normal distribution, standard tests based on normal theory can be used, as given in many statistic books. These tests investigate whether $\therefore$, the mean of X, is the same in each data subset, under the assumption that the variances are the same. For added sophistication, tests of equality of the variances can also be performed:

- To compare the means of two data subsets, perform a Student's t test.
- To simultaneously compare the means of two or more data subsets, perform a one-way analysis of variance test.
- To compare the variances of two data subsets, perform an F test.
- To compare variances of two or more data subsets, use some version of a likelihood ratio test, such as Bartlett's test or a Pearson-Hartley test, as discussed by Bain and Engelhardt (1992, p. 426).

These tests are not considered further here, because they rely heavily on the assumption of normality. This is especially true of the tests later in the list. Most statistical software packages will perform these tests. The analyst must ask whether the assumption of normality is established well enough to justify the use of the tests.

Nonparametric Tests Based on Ranks. For general use when normality or lognormality is not well established, nonparametric tests are preferable. The books by Conover (1999) and Hollander and Wolfe (1999) are excellent summaries of standard tests. As before, let $X = \log(T)$, but do not assume that X has a normal distribution or any other particular distribution. Tests for location assume that various data subsets have distributions that are shifted sideways from each other. The shapes are the same, but the medians may be different. This is the nonparametric analogue of assuming that the distributions are normal with a common variance, but possibly different means. Tests for dispersion assume that the shapes are the same, but possibly with different location and scale parameters. This is the nonparametric analogue of assuming normal distributions with possibly different means and variances.

To test equality of two medians against a shift alternative, use the Wilcoxon-Mann-Whitney test. This test was introduced in Section 6.3.3.2.2. In the present context, let W denote the sum of the ranks of times for the first data subset, when all the times are considered together. The ranks are the same whether or not the logarithmic transformation is performed.

For example, to compare group P to group S in Example 6.12, arrange all 70 times from the two groups in ascending order, and mark the times corresponding to group P. The smallest time from group P is 6 minutes. This has rank 12, because it is preceded by 11 values in group S from 2 to 5 minutes. The other ranks are found similarly. Ties are handled by assigning the average rank to all tied values. The rest of the test was explained in Section 6.3.3.2.2. It is not detailed here, because the test is normally performed by a computer.

To test whether two *or more* data subsets can be pooled, the test of choice is the Kruskal-Wallis test. It tests whether the distribution of T is the same in all the data subsets, against the alternative that the distributions have the same shape but different medians. The test is based on a sum of ranks for each data subset. Those who want details can look in Conover (1999) or Hollander and Wolfe (1999); everyone else can just let the computer do the test.

When the Kruskal-Wallis test is applied to all three groups in the data of Example 6.12, it rejects equality of the distributions with p-value 0.026. This is consistent with the graphical comparison in Figure 6.47 — clear evidence of a difference, though not extreme overwhelming evidence. Based on these analyses, Atwood et al. (1998) dropped group P from the analysis of durations, and combined groups S and T. Group P consists of LOSP durations when the plant remained at power throughout the event. The authors comment on reasons why plant personnel might be very deliberate in restoring offsite power while the plant is still up and running.

To test for equality of dispersion of two data subsets, the rank-like test of Moses is recommended. This requires splitting each data subset into two or more parts, and so is not suitable for very small data sets. See Hollander and Wolfe or documentation of a statistical software package for details for applying this test.

A well-known nonparametric test has not been developed for testing equality of dispersion of more than two data subsets. Therefore, graphical comparisons, such as side-by-side box plots, should be an important part of the analysis.

Nonparametric Test Based on EDFs. A well-known test for comparing two data subsets is the two-sample Kolmogorov-Smirnov test. It is based on comparing the empirical distribution functions for the two data sets. The test statistic is

$$D = \max_t[|\hat{F}(t) - \hat{G}(t)|]$$

where $\hat{F}(t)$ and $\hat{G}(t)$ are the empirical distribution functions from the two data sets. Many software packages can perform this test.

6.6.2.2 No Time Trend

This section will be illustrated by an extension of Example 6.12, taken directly from Atwood et al. (1998).

Based on the above type of analysis of Example 6.12, the LOSP study (Atwood et al. 1998) pooled the data from groups S and T, but excluded group P. That report also combined common-cause pairs of events at multiple units into single site-events (one pair of shutdown events, two pairs of trip events, and two pairs that involved a shutdown reactor and a reactor that tripped). This gave a total of 102 site events instead of the 107 in Example 6.12. They are sorted by event date and listed as Example 6.13. Times are in minutes, and dates are MM/DD/YY.

6.6.2.2.1 Graphical Methods

One natural way to examine the data for a trend is through a scatter plot of the observed values against calendar time. Often, as in Example 6.13, a few large values are outliers. They will determine the scale of the vertical axis. Compared to those large values most of the other values are very small, hugging the horizontal axis. In such a case, the observed values should be transformed, typically by taking logs.

Figure 6.48, from the LOSP study (Atwood et al. 1998), shows a plot of $\log_{10}$(recovery time), for the data of Example 6.13. Visually, any trend in time appears to be very slight. The section below, which considers statistical tests, will re-examine this example.

A potentially more helpful plot is a cumulative plot of recovery time against chronological sequence. The vertical axis shows cumulative recovery time, that is, cumulative duration of LOSP events. No logarithmic transformation is made, because a sum of durations is easy to interpret, but a sum of log(duration) is harder to interpret. Also, logarithms can be negative, so a cumulative plot of logarithms would not necessarily be monotone.

Example 6.13 LOSP recovery times and event dates.

4	04/22/80	3.5	11/21/85	40	02/11/91
106	06/03/80	4	11/21/85	240	03/07/91
62	07/15/80	20	12/12/85	67	03/13/91
120	01/16/81	46	01/01/86	29	03/20/91
14	02/01/81	100	01/28/86	60	03/21/91
15	04/27/81	12	07/19/86	277	04/23/91
10	12/23/81	155	03/05/87	24	06/15/91
29	06/22/82	37	03/21/87	60	06/22/91
17	04/26/83	4	04/04/87	20	06/27/91
11	10/04/83	388	07/14/87	11	07/24/91
163	10/08/83	118	07/23/87	4	10/20/91
240	11/14/83	2	08/17/87	77	01/29/92
97	01/08/84	53	09/11/87	20	03/23/92
90	02/12/84	29	09/16/87	20	03/27/92
15	02/16/84	17	10/14/87	35	04/02/92
2	02/28/84	59	10/16/87	10	04/28/92
5	05/03/84	4	11/17/87	6	05/03/92
2	06/04/84	8	06/24/88	454	08/22/92
120	06/05/84	38	07/17/88	57	10/19/92
20	07/16/84	24	07/29/88	95	12/31/92
11	07/26/84	14	08/13/88	136	04/08/93
10	08/01/84	95	10/16/88	37	05/19/93
20	08/21/84	19	10/25/88	12	06/22/93
22	08/24/84	9	12/26/88	3	06/26/93
3	10/22/84	45	03/25/89	10	09/10/93
14	11/16/84	90	03/29/89	15	09/14/93
15	12/19/84	29	05/14/89	12.5	10/12/93
335	04/29/85	60	06/16/89	96	12/27/93
43	05/07/85	90	06/17/89	2	05/21/94
5	08/16/85	2	06/29/89	1675	11/18/94
73	08/28/85	45	01/16/90	132	02/27/95
25	10/03/85	14	02/26/90	917	10/21/95
13	10/07/85	140	03/20/90	127	01/20/96
60	10/22/85	37	07/09/90	330	02/06/96

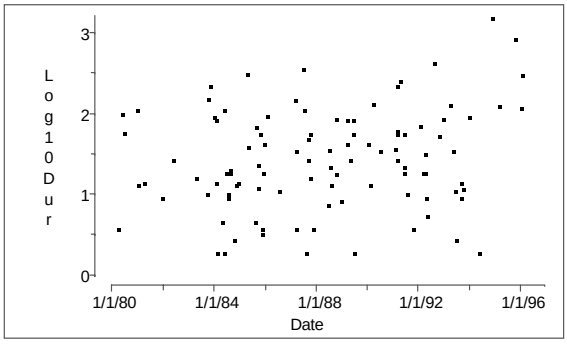


Figure 6.48 Log_{10}(recovery time) plotted against event date, for data from groups S and T in Example 6.13.

If the horizontal axis shows event date, the slope of the curve represents average LOSP duration per calendar time. If, instead, the horizontal axis shows event sequence number, that is, the cumulative number of events, then the slope represents average LOSP duration per event. The latter is more meaningful in a study of durations.

Finally, a diagonal line, connecting the origin to the final point, provides a reference guide, so that the eye can better judge the straightness of the plot.

Figure 6.49 shows the cumulative duration plot for the data of Example 6.13.

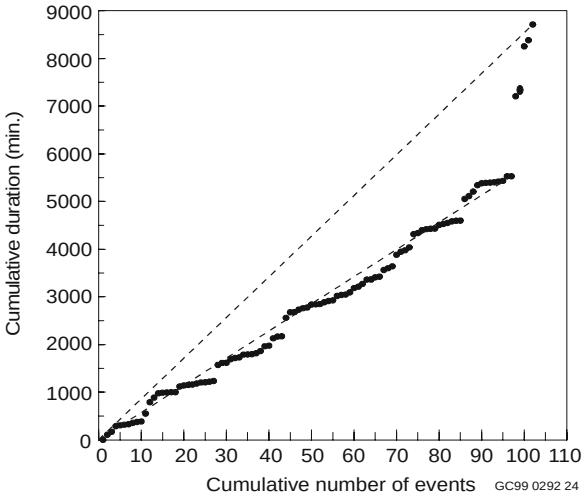


Figure 6.49 Cumulative duration of LOSP events versus cumulative number of events.

The cumulative plot clearly departs from the diagonal straight line, because of two large duration times near the right of the plot. The LOSP report mentions that one of those two times is conservatively large. The LER narrative states that recovery could have been performed earlier, but it does not give an estimated possible recovery time. The LOSP report used times when recovery would have been possible, when such times were available, but for this event the report was forced to use the actual recovery time.

In Figure 6.49, a second dashed line connects the origin (0, 0) to the 97th point, just before the first of the two large jumps. The cumulative plot stays close to this line until the large recovery times occur. Thus, any "trend" is the result, not of a gradual increase in recovery time, but of a couple of outlying values, one of which is conservatively large. Figures 6.48 and 6.49 both reveal the two large recovery times. In this example, however, the cumulative plot seems more informative than the scatter plot, because the log-transformation in Figure 6.48 makes the large times appear less dramatic.

6.6.2.2.2 Statistical Tests

Test Based on Normality. Using the method of least squares fitting, data from a scatter plot may be fitted with a straight line. Most software packages then test of the hypothesis that the slope is zero, assuming normally distributed scatter around the line.

The cited LOSP report fitted a straight line to the data in Figure 6.48 using the least squares method. The trend was reported as statistically significant at the 0.03 level.

This conclusion of a statistically significant trend seems surprising, based on the minimal apparent trend in the figure. The report authors did not have the insights given by the cumulative plot, but they critiqued the calculation in several ways:

- The calculation assumes that log(T) is normally distributed around the trend line. The lognormal distribution (without modeling a trend) was found to fit the data well, and the scatter plot appears consistent with normality. Therefore, the calculated p-value of 0.03 is close to correct.

- The evidence for the trend was very sensitive to the two values in the upper right of the figure. Dropping either value raised the p-value to 0.08. Further, one of those values was known to be conservatively high, as discussed above. This means that the trend may in part be an artifact of the data coding.

- The magnitude of the trend is small. A linear trend in the mean of log(T) corresponds to an exponential trend in the median of T. The magnitude of this trend is a factor of 3.6 over the 17 years of the study. This is fairly small from an engineering viewpoint.

- No solid engineering reasons were found to explain the trend.

Section 6.2.3.1.2 of this handbook discusses how test results should be interpreted. It states that calculation of a p-value is only part of the analysis, and should be followed by critical thinking. The above bulleted list of considerations illustrates that kind of thinking. Use of a cumulative plot would have helped the report authors even more, revealing that a smooth trend of any kind is inappropriate. The authors of the LOSP study chose not to model a trend, but recognized that additional data might change this decision.

Nonparametric Test. A test for a trend that does not assume normality is easy to construct. Such a test is necessary if normality cannot be assumed. If normality can be assumed, the nonparametric test is less powerful for detecting a trend, because it ignores available information, that the data are normally distributed.

The test is the Wilcoxon-Mann-Whitney test, first introduced in Section 6.3.3.2.2. To apply it here, arrange the times sequentially, in order of their event dates. Count an event as *A* if it is above the median and as *B* if it is below the median. Discard any values that equal the median. Now carry out the Wilcoxon-Mann-Whitney test based on the ranks of the *A*s in the sequence of all the events. Because this test is based only on comparisons to the median, it is the same whether or not logarithmic transformations are used.

When this was done with the data from Example 6.13, the median duration was 29. The first duration in Example 6.13 was a *B*, the next three were *A*, and so forth. In all, there were 48 *A*s and 50 *B*s. The *A*s had higher average rank than the *B*s, suggesting an upward trend, but the p-value was 0.09, not quite statistically significant. The nonparametric test is not as sensitive as the parametric test for detecting the small trend, in part because it does not make as much use of the two extreme values seen in Figure 6.49. If the normality assumption were not satisfied, only the nonparametric test would be valid.

6.6.2.3 Goodness of Fit to Parametric Models

One way to model recovery times and other durations is to model the distribution of the durations by some parametric distribution, such as lognormal, Weibull, etc. One must then check to see if the data fit this proposed model well. This section gives graphical methods and statistical tests for such investigations.

6.6.2.3.1 Graphical Methods

The basic idea is to compare nonparametric estimates, which come directly from the data, with estimates based on the fitted model under consideration. For example:

- Compare the histogram to the density from the fitted model.
- Compare the EDF to the c.d.f. of the fitted parametric model. Equivalently, compare the empirical reliability function (1 minus the EDF) to the fitted reliability function.
- Compare the quantiles of the data to the quantiles of the fitted distribution. This plot is called a quantile-quantile plot, or a Q-Q plot. Q-Q plots have become very popular for assessing goodness of fit, although they take getting used to.

These three comparisons are illustrated below, using the data of Example 6.13, and an assumed lognormal distribution. First, the fitted distribution is found by taking natural logarithms of the recovery times, and estimating the mean and variance of their distribution. The estimated mean is 3.389 and the estimated standard deviation is 1.434. The ln(time) values are modeled as normally distributed with this mean and variance. The raw times have the corresponding lognormal distribution.

Figure 6.50 shows the histogram density with a fitted lognormal density overlaid. Because this distribution is concentrated at small values, the goodness of fit is difficult to judge. Therefore, the histogram of the ln(time) values are also plotted, with a normal density overlaid, in Figure 6.51. Actually, the area under the histogram equals the number of observations, and the density has been rescaled to have the same area.

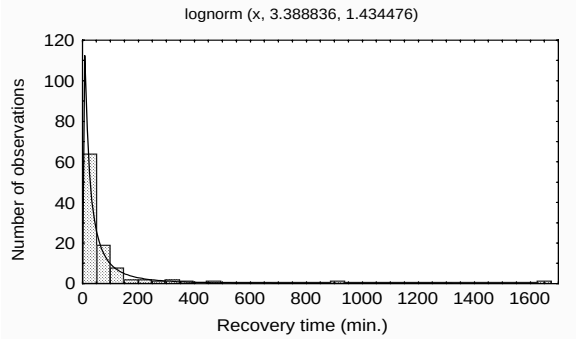


Figure 6.50 Histogram of data from Table 6.19, with multiple of lognormal density overlaid. The skewness makes goodness of fit difficult to assess.

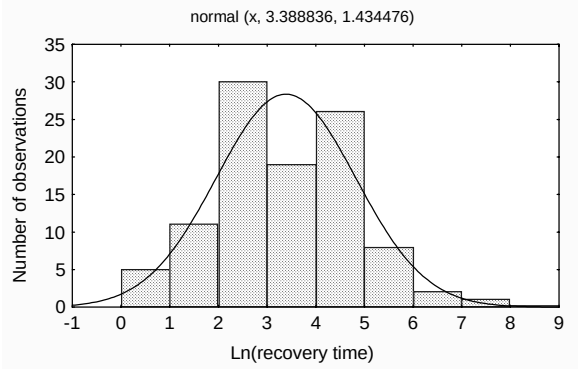


Figure 6.51 Histogram of ln(time), with a multiple of a normal density overlaid. Fit appears as good as achievable without using a bimodal distribution.

Figure 6.52, from the LOSP report, shows a plot of the reliability function, 1 minus the EDF, with the corresponding fitted function, 1 minus the lognormal c.d.f. The plot in this form is useful for interpreting the degree to which the fitted c.d.f. differs from the empirical c.d.f., because the horizontal axis is in units of time. A plot in terms of log(time) would not hug the axes so closely. Therefore, discrepancies

between the curves would be more visible, but their magnitudes would be harder to interpret in real-world terms.

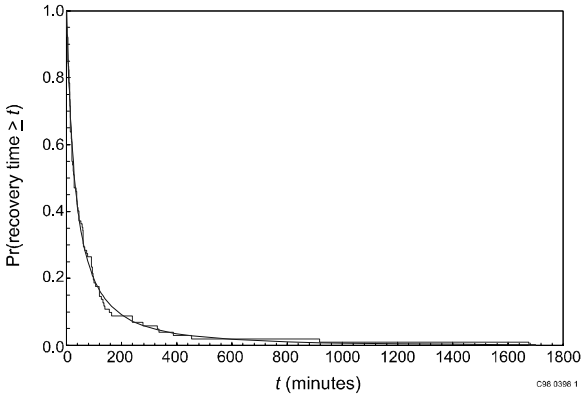


Figure 6.52 Empirical and theoretical reliability functions, where the reliability function is defined as 1 minus the c.d.f.

Finally, Figure 6.53 gives a quantile-quantile (Q-Q) plot, described by Wilk and Gnanadesikan (1968). If only one plot could be used, a Q-Q plot would be a strong contender. A Q-Q plot compares two distributions by plotting the quantiles of one against the corresponding quantiles of the other. If X is a linear function of Y, $X = a + bY$, then a Q-Q plot of X versus Y will be linear. The parameters a and b do not need to be known or estimated; linearity of the plot tells the analyst whether the two distributions are the same except for a linear transformation. Users of probability paper will recognize that a plot on probability paper is a form of a Q-Q plot.

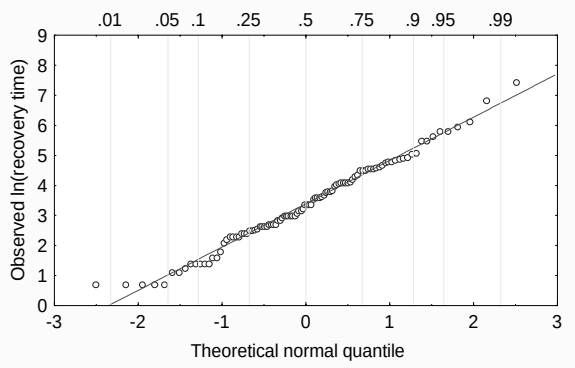


Figure 6.53 Quantile-quantile plot of ln(recovery time) and fitted normal distribution. The points fall nearly on a straight line, indicating good fit.

In Figure 6.53, the software package implemented the Q-Q plot by plotting the ordered values of ln(time) against the theoretical expected values of the corresponding order statistics assuming normality. For example, denote ln(t) by y. In the implementation of this particular software package, the ith ordered value, $y_{(i)}$, is plotted against the expected value of $Z_{(i)}$, assuming that 102 values of Z are randomly sampled from a standard normal distribution. The points follow a straight line. This indicates that the data are, apparently, normally distributed.

The parameters, μ and σ, can be ignored in a Q-Q plot based on the normal distribution, because a normal random variable Y with mean μ and standard deviation σ is related to Z by $Y = \mu + \sigma Z$. This is a linear transformation, and so does not change the linearity or nonlinearity of the plot. In fact, it is not even necessary to obtain estimates of μ and σ. For distributions other than normal, the parameters may need to be estimated before the Q-Q plot can be constructed.

The expected values of the order statistics cannot be constructed without tables or a computer program. Users of probability paper may construct a simpler version, plotting $y_{(i)}$ against the $i/(n+1)$ quantile of a standard normal distribution. Here n is the total number of observations, 102 in the present example. This simpler version gave its name to the plot, a quantile-quantile plot.

For the purpose of illustration, Figure 6.54 gives a Q-Q plot of the same example data, assuming that the raw recovery times have a *normal* distribution. Of course the fit is horrible — no one expects the raw times to have a normal distribution. This lack of fit is shown by strong curvature in the plot. The two largest times show the lack of fit most emphatically, but even without them the plot appears to show a curvature that indicates non-normality.

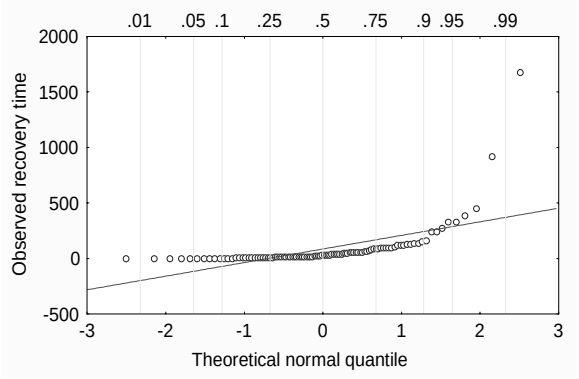


Figure 6.54 Quantile-quantile plot of raw recovery times against fitted normal distribution. The strong curvature indicates bad fit.

The particular form of the distribution can sometimes allow special tricks. Let us leave the present example, and consider investigating whether data t_1, ..., t_n come from an exponential distribution. Example 6.6, which was deferred from Section 6.2.3.4, will be used to illustrate the method.

The idea of the Q-Q plot is that, when the data come from the assumed distribution, then

$t_{(i)} \approx F^{-1}[i/(n+1)]$,

where F^{-1} is the inverse of the assumed c.d.f. Let us find the inverse of the exponential c.d.f. Set

$y = F(t) = 1 - e^{-\theta t}$.

To find the inverse, solve for $t = F^{-1}(y)$:

$e^{-\theta t} = 1 - y$

and therefore

$t = -\ln(1 - y)/\theta$.

The right-hand side is $F^{-1}(y)$, so the defining relation of the Q-Q plot is

$t_{(i)} \approx -\ln[1 - i/(n+1)]/\theta$.

Thus, a plot of the ordered times against $-\ln[1 - i/(n+1)]$ should be approximately linear, regardless of the value of θ. The linearity or nonlinearity of the plot does not depend on whether θ has been estimated well. Nonlinearity is evidence against the assumed exponential distribution.

Example 6.6 contains times between LOSP events, which should be exponentially distributed. A plot of the ordered times against $-\ln[1 - i/(n+1)]$ is shown in Figure 6.55. Because the plot does not show much curvature, it indicates good fit to the exponential distribution.

6.6.2.3.2 Statistical Tests

The previous section used graphs to investigate whether data followed a certain kind of distribution. The present section gives statistical tests of hypotheses, for investigating the same question. The tests here are called **goodness-of-fit tests**, because they are intended to test whether the data fit the assumed model well. The null hypothesis is that the data come from a distribution of the assumed form, for example, from a lognormal distribution. The null hypothesis does not specify the parameters. Therefore, the null hypothesis includes a family of distributions. The alternative hypothesis is that the data come from some other distribution.

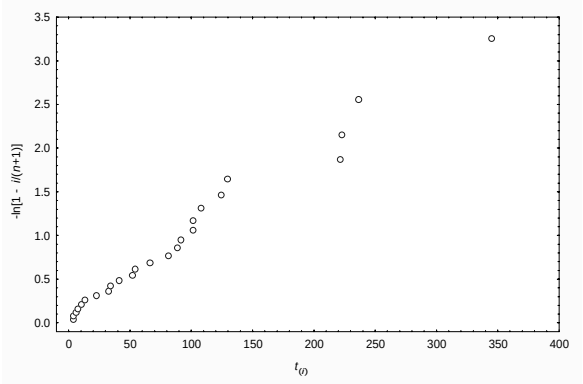


Figure 6.55 Q-Q plot for checking exponential distribution in Example 6.6.

As always, remember that "acceptance" of the null hypothesis does not provide evidence that the null hypothesis is true. It merely indicates a lack of evidence that the null hypothesis is false. For example, the data may be consistent with a lognormal distribution, and also consistent with a gamma distribution and a Weibull distribution. In such a case, the analyst should not make assertions that are highly dependent on the form of the distribution. For example, a sample of 10 observations may be consistent with many possible distributions. An estimate of the 99.9th percentile of the distribution would be a large extrapolation from the actual data, highly dependent on the assumed form of the distribution. A confidence interval on this percentile would be even worse, because it would give an appearance of quantified precision, when in reality the distribution could have practically any form out in the tail.

In summary, even though a model has been "accepted," it is only an approximation. The analyst should not make assertions that are sensitive to small departures from the model.

Chi-Squared Test. The chi-squared test, seen in Sections 6.2 and 6.3, is also an all-purpose goodness-of-fit test. To apply it in the present context, estimate any unknown parameters of the hypothesized distribution of T. Based on these parameter estimates, divide the time axis into c bins of equal probability. The letter c stands for *cell*, another term for a bin in this context. Based on the recommendations of Moore (1986), choose the number of bins so that n/c is at least 1, and preferably at least 2. Let x_i be the observed number of

values of T in the ith bin. Because the bins have equal probability, the expected number of values of T that will fall in any bin is n/c, the number of observations divided by the number of bins. The Pearson chi-squared statistic is

$$X^2 = \sum_j (x_j - e_j)^2/e_j \quad ,$$

where each e_j equals n/c and each x_j is an observed count.

If the null hypothesis is true, the distribution of X^2 is approximately chi-squared. The commonly quoted rule is that the degrees of freedom is $c - 1 - p$, where p is the number of estimated parameters. For example, suppose the null hypothesis is that the distribution of T is lognormal, or equivalently, that ln(T) is normal. Then two parameters must be estimated, μ and σ. Thus, the commonly quoted rule for the degrees of freedom is $c - 3$. In fact, researchers have found that this is not quite correct, for subtle reasons described by Moore (1986, Section 3.2.2.1). The correct degrees of freedom are somewhere between $c - 1 - p$ and $c - 1$. The exact value depends on the form of the distribution in the null hypothesis.

Let us apply this to the LOSP-recovery data from Example 6.13, and use X = ln(T) for convenience. Let H_0 be the hypothesis that X is normally distributed. As mentioned above, the estimates of μ and σ are 3.389 and 1.434. With 102 observations, it is convenient to take 50 bins, so that each expected count is 102/50 = 2.04. The bin boundaries are the 0.02, 0.04, ..., 0.98 quantiles of the distribution. These are estimated as

$$y_q = 3.389 + 1.434 z_q \, ,$$

where q is 0.02, 0.04, etc., and z_q is a quantile interpolated from a table of the standard normal distribution. For example, $z_{0.02} = -2.054$.

When this is carried out, using a computer to perform the calculations, the value of X^2 is 63.69. The distribution under H_0 is chi-squared with degrees of freedom between 47 and 49. Therefore, the p-value is between 0.053 and 0.077. The test almost rejects normality of ln(T) at the 0.05 level, in spite of the graphical evidence to the contrary.

Upon examination, the test is revealed to be too powerful for its own good. It notices that the values tend to cluster, five occurrences of 2 minutes, six values of 20 minutes but no values of 21 minutes, etc. With 50 cells, each observed time is commonly the sole occupant of a cell. The test notices that the numbers have been rounded to convenient times, such as 20 minutes, and uses this as evidence against normality. In fact, such clustering is a departure from normality, and from any other continuous distribution. But it is not the kind of departure that is of interest to most analysts.

A coarser binning, into fewer cells, would not be distracted by fine clustering, and would search for more global departures from the null hypothesis.

We conclude this discussion of the chi-squared test by considering again the exponential example that was deferred from Section 6.2.3.4.

Example 6.6 consists of 25 times. The null hypothesis is that the data come from an exponential distribution. The unknown θ is estimated as the number of events divided by the total observation period, 25/(2192 days) = 0.0114 events per day. This MLE is justified based on the Poisson count of events, as in Section 6.2.1.1. To obtain a moderate expected count in each bin, let us use ten bins. They have equal estimated probabilities, 0.10 each, if they run from

0 days to [-ln(0.9)]/0.0114 = 9.24 days
9.24 days to [-ln(0.8)]/0.0114 = 19.57 days
...
201.89 days to infinity.

These calculations are all based on the exponential c.d.f., $F(t) = 1 - \exp(-\theta t)$. Setting $F(t)$ to 0.1, 0.2, and so forth gives the bin boundaries.

There are four observed times in the first bin, two in the second, and so on. The expected count in each bin is 25/10 = 2.5. The calculated value of X^2 is 9.00. This must be compared with the percentiles of the chi-squared distribution. There are c = 10 bins, and p = 1 estimated parameter. Therefore, the degrees of freedom are between 10 - 1 = 9 and 10 - 2 = 8. The value 9.00 is in the middle of both of these distributions, the 56th percentile of one and the 66th percentile of the other. Therefore, the chi-squared test finds no evidence against the exponential distribution. This agrees with the earlier graphical analysis.

Shapiro-Wilk Test for Normality. Many software packages offer the Shapiro-Wilk test for normality. It is based on obseving how closely the order statistics follow theoretical normal values, as displayed for example in Figure 6.53. For testing the normal distribution, the Shapiro-Wilk test is one of the most powerful tests against a wide variety of alternatives. Details are not given here, because all the calculations are carried out by the computer.

With the logarithms of the data of Example 6.13, the Shapiro-Wilk test does not reject normality of ln(T), giving a p-value of 0.34. This agrees with the visual evidence of Figure 6.53.

Tests Based on the EDF. Several families of tests have been proposed based on the empirical distribution function (EDF, defined in Section 6.6.1.1.3). The idea is to reject the null hypothesis if the EDF is not "close to" the theoretical c.d.f. Closeness can be measured in various ways, giving rise to a variety of tests. EDF-based tests are appealing because they do not require a choice of bins, but simply use the data as they come.

The most famous such test is the Kolmogorov test, also known as the Kolmogorov-Smirnov test. It is described in Appendix B.3.4. This test differs from the similarly-named test in Section 6.6.2.1.2 because the present test asks whether a random variable has a certain distribution, and the earlier test asks if two random variables have the same distribution. These are slightly different questions. The test here rejects H_0 if

$$\max |\hat{F}(t) - F(t)|$$

is large, where the maximum is over all values of t. Here, any unknown parameters in F must be estimated; the effect of this estimation is typically ignored.

When SAS (SAS Version 8, 2000) performs the Kolmogorov test of lognormality on the times in Example 6.13, it gives a p-value > 0.15. That is, it does not calculate the exact p-value, but it does report that the departure from lognormality is not statistically significant.

The Cramér-von Mises test and the Anderson-Darling test are other EDF-based tests, designed to remedy perceived weaknesses in the Kolmogorov test. The Cramér-von Mises test is based on

$$\int [\hat{F}(t) - F(t)]^2 f(t) dt .$$

Here, F is the distribution that is assumed under the null hypothesis, and f is the corresponding density. Thus, the Kolmogorov test looks at the maximum difference between $\hat{F}$ and F, while the Cramér-von Mises test looks at an average squared difference. The Anderson-Darling test is based on

$$\int \left\{ [\hat{F}(t) - F(t)]^2 / \{F(t)[1 - F(t)]\} \right\} dt .$$

This division by $F(t)[1 - F(t)]$ gives greater weight to the tails of the distribution, where departures from F is most likely to occur. Thus, this test is intended to be more powerful than the Cramér-von Mises test against common alternative hypotheses. Many statistical packages perform one or more of these tests.

When testing lognormality of the data in Example 6.13, SAS reports a p-value of >0.25 for the Cramér-von Mises test and also for the Anderson-Darling test. Just as for the Kolmogorov test, SAS does not compute the exact p-value, but it does report that the departure from lognormality is not statistically significant.

6.6.2.4 Consistency of Data with Prior in Bayesian Parametric Estimation

The issue here is whether the data are consistent with an assumed *informative* prior distribution for the unknown parameters. If a noninformative prior distribution is used, then the question does not arise, because the noninformative distribution is supposed to be consistent with anything.

6.6.2.4.1 Exponential Durations

A quantitative approach is possible when T has an exponential(θ) distribution. In this case all the information of interest about θ is contained in $G t_i$, as seen in Section 6.6.1.2.2. Therefore, we can compare $G t_i$ to what would be expected based on prior belief about θ.

If $G t_i$ is surprisingly large or surprisingly small, that is, if $G t_i$ is in either tail of the distribution of $G T_i$, then the prior distribution is questionable. The value $G t_i$ is in the lower tail if $\Pr(G T_i < G t_i)$ is a small probability, and in the upper tail if $\Pr(G T_i > G t_i)$ is a small. To be specific, consider the upper tail. The relevant probability is

$$\Pr(G T_i > G t_i) = \int \Pr(G T_i > G t_i \mid \theta) \, f_{\text{prior}}(\theta) \, d\theta . \quad (6.25)$$

The inner conditional probability can be evaluated by using the fact that the distribution of $G T_i$, given θ, is gamma(n, θ). If the prior distribution of θ is not conjugate, the integral in Equation 6.25 must be evaluated numerically, just as in Sections 6.2.3.5 and 6.3.3.4: either (a) compute the integral using numerical integration, or (b) generate a random sample of θ values from the prior distribution, find $\Pr(G T_i > G t_i \mid \theta)$ for each such θ, and find the average of these probabilities as the overall probability.

Treatment of the lower tail follows the same pattern.

If the prior distribution of θ is conjugate, that is, gamma(α, β), then Equation 6.25 simplifies. By working out the integrals it can be shown that $G T_i/(G T_i + \beta)$ has a beta(n, α) distribution. Equivalently, $\beta/(G T_i + \beta)$ has a beta(α, n) distribution. These are marginal distributions corresponding to Equation 6.25, from which θ has been integrated out. Therefore, if $G t_i /(G t_i + \beta)$ is in either extreme tail of a beta(n, α) distribution, or equivalently, if $\beta/(G t_i + \beta)$ is in either extreme tail of a beta(α, n) distribution, then the gamma(α, β) prior distribution is questioned.

In Example 6.12, suppose that the only events of interest are those in the group of S (shutdown) events. Suppose also that the times are assumed to be exponential(θ) – the realism of that assumption is not the subject of the present investigation. Finally, suppose that θ is assigned a gamma(2, 30) prior distribution, roughly equivalent to two prior observed times with total duration of 30 minutes. The shape parameter of only 2 means that the prior is not very informative, so we expect the data to be consistent with it, unless 30 minutes is very unrealistic.

From Table 6.15, we find n = 62 and the total of the durations is 62×92.3 = 5722.6. The beta tables in Appendix C assume that the first beta parameter is smaller than the second, so it is convenient to work with the beta(2, 62) distribution rather than the beta(62, 2) distribution. Therefore, we ask if

30/(5722.6 + 30) = 5.2E!3

is in either tail of a beta(2, 62) distribution. Table C.5 shows that the 5th percentile of the beta(2, 62) distribution is roughly 6E!3 (it is an interpolation of 7.01E!3 and 3.53E!3 in the table). Table C.6 shows that the 2.5th percentile is roughly 4E!3. So the observed value is somewhere between the 2.5th and 5th percentiles of the predictive distribution. This means that the prior may need rethinking. It should either be modified or it should be justified more carefully. (In the present example the prior came out of thin air, but the real difficulty is that the durations are not really exponential – the whole exercise is only for illustration.)

6.6.2.4.2 Distributions Having Two or More Parameters

When the topic of comparing the data to the prior arose in connection with estimating θ or p, there was a single parameter of interest, and a single observed random variable that contained all the information of interest for that parameter. This random variable was the total count of initiating events, the count of failures on demand, or, in the previous section, the total duration. However, the present subsection considers a distribution with (at least) two parameters, such as λ and Γ or α and β. No single random variable contains all the information of interest. Therefore, in such cases it is simplest to compare the data with the prior by constructing:

1. a prior credible region for the two parameters, and
2. a posterior credible region *based on noninformative priors*.

The first case shows what the prior distribution says, and the second case shows what the data say. Compare the answers from 1 and 2 to see if the prior distribution and the data seem consistent, that is, if the prior region contains most of the posterior region. Instead of two-dimensional credible regions, one might calculate credible intervals for the individual parameters. This is simpler, but ignores the possible correlation of the two parameters. Because this is such an advanced topic, no examples are worked out here.

6.6.3 Nonparametric Density Estimation

The most prevalent methods of estimating a density function are parametric methods. As described in Section 6.6.1.2, the density is specified in terms of a functional form, such as lognormal or Weibull, with unknown parameters. The parameters are then estimated from the data. However, there also exist nonparametric methods for estimation of a density function, some of which are described here.

The simplest and best known method of estimating a density function is to construct a frequency table, and then to plot the histogram. This method was discussed in Section 6.6.1.1.4. Two illustrations are given there, Figures 6.43 and 6.44. Both use the 45 recovery times from part T of Example 6.12. The methods discussed below are illustrated with the same set of 45 recovery times.

6.6.3.1 Smoothing Techniques and Kernel Estimators

Smoothing techniques can be motivated by recalling that the density function, $f(t)$, is the derivative of the c.d.f., $F(t)$. The EDF, discussed in Section 6.6.1.1.3 and denoted by $\hat{F}(t)$, is a natural estimator of $F(t)$. Thus, a natural estimator of the density is the differential quotient using the EDF in place of the c.d.f.,

$$\hat{f}_n(t) = \frac{\hat{F}(t+h) - \hat{F}(t-h)}{2h}, \qquad (6.26)$$

where h is an increment of the variable t. The main problem in applying such an estimator is to choose h small enough so that the differential quotient adequately approximates the derivative, but large enough so that the interval with limits $t \pm h$ contains a sufficient amount of data.

Recall that $\hat{F}(t)$ equals the number of observations having a value less than or equal to t divided by the total number of observations, n. Therefore, Equation 6.26 can also be written as

$$\hat{f}_n(t) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{t - t_i}{h} \right) , \qquad (6.27)$$

where K is a function defined as $K(u) = 1/2$ if u is between ± 1, and zero otherwise, and t_i is the ith observation. Notice that an observation t_i only enters into this calculation if $(t_i \; ! \; t)/h$ is between ± 1, or in other words if t_i is near t; specifically if t_i is within h units of t. Thus, the estimate is based on averaging values of 1/2 when observations are near t. This is a special case of a general type of estimator known as a **kernel density estimator**. The function $K(u)$ is called the kernel and the increment h is called the **bandwidth.** The bandwidth defines a "window", centered at t and having width $2h$, which contains the data involved in the estimate at the point t.

6.6.3.1.1 The Rectangular Kernel

When graphed, the kernel corresponding to Equation 6.27 is a rectangle of height 1/2 and width $2h$. The resulting estimator is illustrated here with group T of Example 6.12 and two bandwidths.

Figure 6.56 shows a graph of the estimate of the density when the bandwidth is h = 25 minutes. Notice that the estimated density is zero in the interval roughly from 150 to 250 minutes. This corresponds to the fourth and fifth bins of the histogram of Figure 6.43, both of which were empty.

It is also evident that the graph is somewhat jagged, indicating that the bandwidth may be so small that not enough data are being captured in the window.

The vertical dashed line marks the point t = 0, to be discussed later.

Consider now a rectangular kernel estimate with the same data but with a larger bandwidth, h = 50 minutes. The results are shown in Figure 6.57.

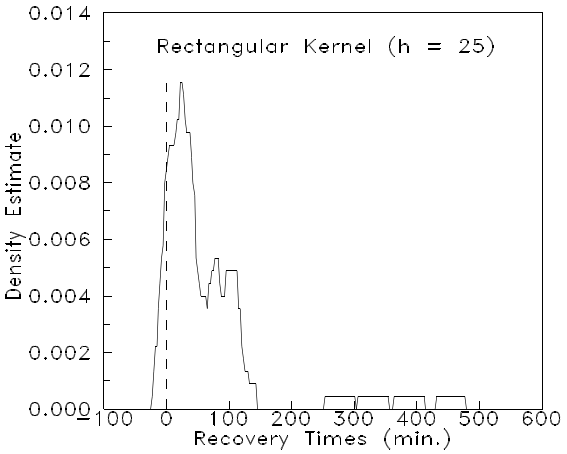


Figure 6.56 Density estimate of the data from group T in Example 6.12, with rectangular kernel and bandwidth 25.

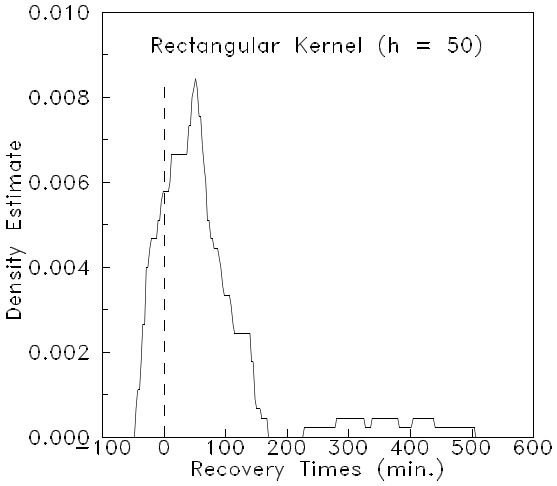


Figure 6.57 Density estimate of the data from group T in Example 6.12 with rectangular kernel and bandwidth 50.

There is still some jaggedness, but it is somewhat less than in Figure 6.56. There is still a noticeable low point in the vicinity of 200 minutes, but it is narrower than in Figure 6.56.

It is clear that by smoothing over a very wide window, any features can be smoothed out. For this reason, it is desirable to give some thought to whether there is some explanation for low density. In other words, are these real effects or are they just due to randomness? If the low estimates can be explained by something other than random fluctuation, smoothing would tend to hide this fact, but if they are due to randomness, then smoothing should be helpful.

This issue was also seen with histograms. Choosing too narrow bins for the size of the data set caused the shape to be influenced too much by random variation. Choosing too wide bins smoothed out nearly all the variation. The question of how much to smooth and how much roughness to allow is inherent in all forms of density estimation.

6.6.3.1.2 Boundary Problems

Notice, that as the bandwidth is increased, the interval over which the estimated density is positive becomes wider. This is because the window is picking up more data as it gets wider. This causes the anomaly that the estimated density is positive over negative values of the t axis, even though t represents a positive variable, namely recovery time. The vertical dashed line marks the point $t = 0$ in each figure, and the portion of the density to the left of this line is substantial. In addition, although many values of t_i are close to zero, the density estimate *decreases* as t moves leftward to zero. Various methods have been suggested for correcting the estimate at the boundary of the possible region.

Silverman (1986) gives a method that is very easy to implement. If the density is allowed to be positive only for $t \geq 0$, augment the data by reflecting it around 0. That is, create a new data set that consists of

$$\{..., -t_2, -t_1, t_1, t_2, ... \} .$$

Estimate the density based on this data set. Call this estimate $\tilde{f}(t)$. The integral from -4 to 4 of $\tilde{f}(t)$ is 1.0, because $\tilde{f}$ is a density. Also, if the kernel is a symmetrical function, then $\tilde{f}$ is symmetrical around zero, that is, $\tilde{f}(-t) = \tilde{f}(t)$. Now, define the real density estimate by

$$\hat{f}(t) = 0 \quad \text{for } t < 0$$

$$\hat{f}(t) = 2\tilde{f}(t) \quad \text{for } t \geq 0 .$$

Then, $\hat{f}$ is a density that is zero for negative t and nonnegative for positive t. It estimates the unknown true density.

Figure 6.58 shows the resulting estimate with the data of this section, when the kernel is rectangular and the bandwidth $h = 50$. This estimate can be compared with Figure 6.57.

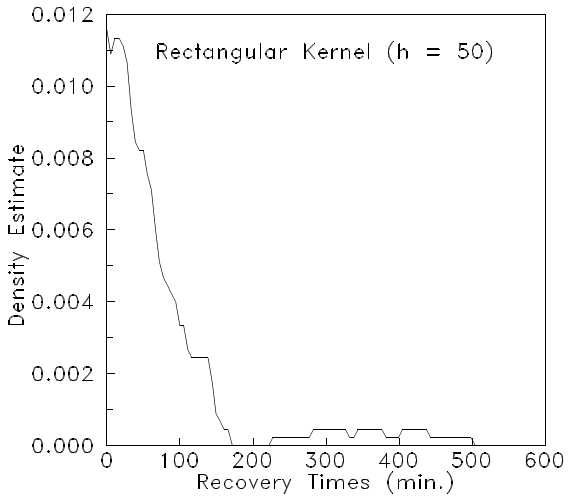


Figure 6.58 Density estimate from group T of Example 6.13, with rectangular kernel and bandwidth 50, forced to be nonzero on positive axis only.

For large t, this estimate is very similar to that of Figure 6.57. However, it is quite different for t near zero. The density is not plotted for $t < 0$, but it equals zero there.

The simple method just given forces the density estimate to have slope zero at the boundary. Those who want to allow a density estimate with nonzero slope at the boundary can see Hart (1997, Sec. 2.5). Technically, Hart's book deals with smoothing a scatter plot, but the method given there can be adapted as follows to smoothing a density estimate: construct a rough histogram density estimate, place a dot at the top of each histogram bar (including the bars with height zero!), and treat those dots as a scatter plot.

6.6.3.1.3 The Triangular Kernel

It may also be desirable in some cases to give less weight to the data in the extremes of the window and to produce a smoother graph. This can be accomplished by choosing a different function for the kernel. A very simple one which does this is the function $K(u) = 1 - |u|$ if u is between ± 1, and zero otherwise. The graph of $K(u)$ is an isosceles triangle with a base two units in width. This kernel gives more weight to the data in the middle of the window and less to data at the sides of the window. It is also possible, by choosing a kernel function with a smoother graph, to produce a kernel estimate which is also smoother. The normal kernel, given next, is such a smooth kernel.

6.6.3.1.4 The Standard Normal Kernel

A kernel function that is often used is the **standard normal kernel**, equal to the standard normal p.d.f., which is given in Appendix A.7.2.

Figure 6.59 shows the density estimate for the same recovery time data, but using the standard normal kernel and bandwidth 25. The density has been made positive on the positive time axis only, by the technique of Section 6.6.3.1.2.

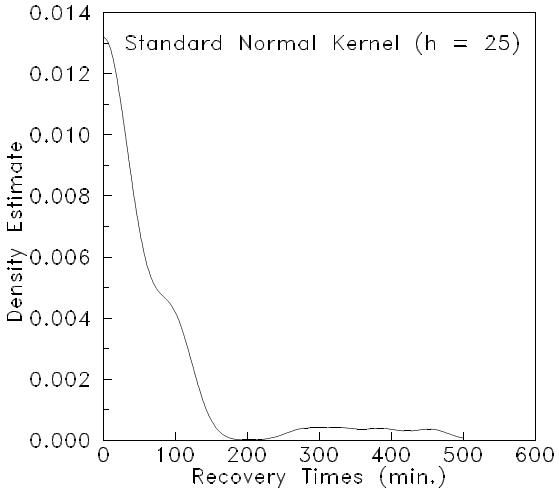


Figure 6.59 Density estimate of the data from group T in Example 6.12, with standard normal kernel and bandwidth 25.

The resulting plot is clearly much smoother than the ones obtained using the rectangular kernel. The increased smoothness is provided by the standard normal kernel, which is differentiable everywhere. The low estimate of density near 200 is still present, but the low spot does not drop to zero as it did in Figure 6.58. This is because the standard normal kernel is always positive. Even though this kernel gives less weight to data which are farther from the center of the kernel, it makes use of every observation in the data set. Consequently, with the standard normal kernel, all terms in the density estimate of Equation 6.27 are positive, although the extreme ones will tend to be relatively small.

For the sake of comparison, Figure 6.60 shows the standard normal kernel estimates for bandwidth h = 50.

Although the graphs shown in Figures 6.59 and 6.60 retain some general features of the graphs in Figures 6.56 through 6.58, they are somewhat smoother. As mentioned in the case of the rectangular kernel in Section 6.6.3.1.1, this type of smoothing is desirable if the sparsity of data in these intervals is due to randomness, but possibly undesirable if there is an explanation for the sparseness.

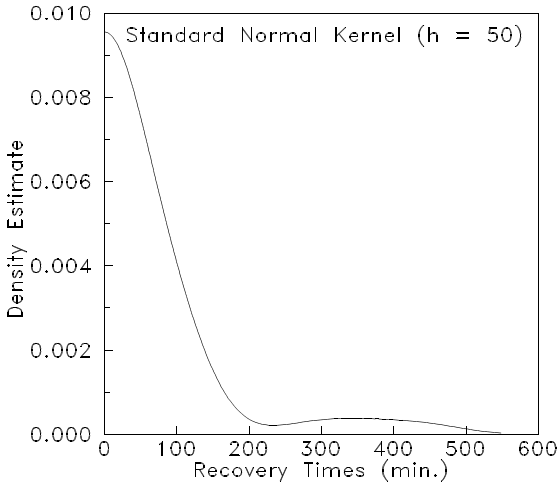


Figure 6.60 Density estimate of the data from group T in Example 6.12, with standard normal kernel and bandwidth 50.

6.6.3.2 Choosing the Bandwidth

General guidelines for choosing a kernel and bandwidth are difficult to formulate. The choice of a bandwidth always involves a trade-off between bias and variability. An attempt to reduce bias generally requires a small bandwidth, but this tends to result in a large variance. On the other hand, choosing a large bandwidth will reduce the variance, but at the expense of increasing the bias. A criterion which accounts for both the bias and variance is based on a quantity called the **mean squared error** (MSE) which is equal to the mean squared difference between the unknown parameter and its estimator. It is easy to show that

$$\text{MSE} = (\text{bias})^2 + \text{variance of estimator}$$

so that as the MSE approaches zero, both the bias and the variance of the estimator also approach zero.

A reasonable choice of bandwidth should take into account the amount of data, and so the solution must depend on n. Thus, we consider a sequence, $h = h(n)$. The sequence should converge to zero, but not too quickly or too slowly. It is known, for example, that under certain fairly modest assumptions, a desirable form for the bandwidth is

$$h(n) = cn^{-1/5}.$$

The main problem is that calculation of the constant c requires more than is typically known about the p.d.f. to be estimated, and it also depends on the choice of a kernel. For example, according to p. 45 of Silverman (1986), for the standard normal kernel, and assuming the distribution of the data to be normal with standard deviation σ, the bandwidth which minimizes the integrated MSE asymptotically is

$$h(n) = 1.06\sigma n^{-1/5}.$$

Notice that the constant c in this case requires that the standard deviation be known or at least estimated.

For example, with the recovery time data, the sample standard deviation, given in Table 6.16, is 99.9 minutes. If this is used to estimate σ, then the optimal bandwidth is $h(n) = 105.9n^{-1/5}$. Using the sample size $n = 45$ yields $h = 49.5$. This result is very nearly the bandwidth of 50 minutes that was used in Figure 6.60.

Keep in mind that this choice of bandwidth was derived for the case where both the distribution being estimated and the kernel are normal, so the result would be good with these assumptions. However, this might be a good place to start if trial and error is used to determine what bandwidth to use. In other words, if it is not clear what to assume, then it would be best to try a few different bandwidths and choose one which provides some smoothing, but does not obscure basic features. As Silverman says, "There is a case for undersmoothing somewhat; the reader can do further smoothing 'by eye' but cannot easily unsmooth."

Another problem that often occurs in practice is that the data will be plentiful in some parts of the range, but sparse in others. This is typical with data from highly skewed distributions. For example, with a positively skewed distribution, such as any of the distributions in Section 6.6.1.2, there will tend to be more data in the lower end than in the upper tail. This would suggest the desirability of having a bandwidth that varies with t, so that a shorter increment can be used in the lower end where the data points are more abundant, and a larger increment used in the upper tail where there are not as many points. This idea is not developed here, but such methods exist. For additional reading on this topic, see the discussions of the nearest neighbor method and the variable kernel method in Silverman (1986).

6.7 Unavailability

Data analysis methods for unavailability do not have the long history that they have for other parameters. Most of the material here is taken directly from a paper by Atwood and Engelhardt (2003). Users of the methods described below should refer to the final peer-reviewed and published version of the article.

As in Section 2.6, the discussion here is presented in terms of trains, although other hardware configurations, such as individual components, could be considered. The terminology of Section 2.6 will be used: outage times are random and the unavailability is a parameter, an unknown constant, denoted here by q. Subscripts, such as "planned" and "unplanned," can be attached to q for clarity if needed.

Two possible kinds of data are considered:

- **Detailed data**: the onset time and duration of each individual outage are recorded, as well as the total time when the train should have been available.
- **Summary data**: the history is grouped into reporting periods, such as calendar months, and the total outage time and total exposure time are recorded for each reporting period.

The methods of this section are illustrated with Example 2.16, taken from Atwood and Engelhardt (2003). This example concerns the unavailability of a train in a two-train CVC system. The 21 outages are given in Table 6.17, and monthly summary data are shown in Table 6.18. All times are in hours.

6.7.1 Analysis of Detailed Data

Denote the mean duration of an outage by **MTTR**, an acronym for **mean time to repair**. Similarly, denote the mean duration of the period between outages (the up time) by **MTTF**, an acronym for **mean time to failure**. Both of these can be estimated when the individual outage times and exposure times are recorded.

Under the model assumptions given in Section 2.6.2, it can be shown (see Ross 1983, pp. 66-67) that the unavailability equals

$$q = \frac{MTTR}{MTTF + MTTR}. \qquad (6.28)$$

This provides a basis for estimating availability and unavailability when the full duration data are recorded.

Upon reflection, this striking result is not surprising. If the data consist of n up times, with n down times interspersed, MTTR would be estimated by (total down time)/n, and MTTF would be estimated by (total up time)/n. Based on Equation 6.28, the natural estimate of unavailability would be

[down-time/n]/[up-time/n + down-time/n]
= down-time/(up-time + down-time) ,

which is just the observed fraction of time when the system is down.

Table 6.17 Detailed data for Example 2.16.

Month	Train Exposure	Train 1 Outages	Train 2 Outages
1	364	0	0
2	720	25.23	24.88 75.08
3	744	0	0
4	711	23.45	0
5	621	9.75 0.49 1.24	15.15 1.02 0.49
6	502	0.34 2.90 4.43	4.43 0.34
7	0	0	0
8	637	18.02	12.47 9.48 18.02
9	676	0	0
10	595	0	0
11	600	11.05	0
12	546	0	0
13	745	0	52.25
14	720	0	0
15	744	0	0
Total	8,925	96.90	213.61

In addition, MTTF + MTTR is the mean time from one outage onset to the next, which can be interpreted as the reciprocal of the outage frequency. Therefore, the unavailability can be rewritten as

q = (outage frequency)
× (mean outage duration) . (6.29)

This equation is the foundation of all the methods for using detailed data to estimate q.

6.7.1.1 Frequentist Point Estimate

Suppose that n outages are observed in exposure time t_{expos}, with total outage duration t_{dur}. Based on Equation 6.29, the simple frequentist point estimate of q is

$(n/t_{expos})\times(t_{dur}/n) = t_{dur}/t_{expos}$.

Table 6.18 Summary data for Example 2.16.

Month	Train Exposure	Train 1 Outage Time	Train 2 Outage Time	Total Outage Time
1	364	0	0	0
2	720	25.23	99.96	125.19
3	744	0	0	0
4	711	23.45	0	23.45
5	621	11.48	16.66	28.14
6	502	7.67	4.77	12.44
7	0	0	0	0
8	637	18.02	39.97	57.99
9	676	0	0	0
10	595	0	0	0
11	600	11.05	0	11.05
12	546	0	0	0
13	745	0	52.25	52.25
14	720	0	0	0
15	744	0	0	0
Total	8,925	96.90	213.61	310.51

This is the moment estimate, replacing the means in Equation 6.28 by the sample means. It is a natural, intuitively appealing point estimate.

For the data of Example 2.16, either Table 6.17 or 6.18 shows that the estimate of q is 310.51/17850 = 1.74E!2. (Note, the exposure time of 8925 hours for a single train must be doubled to get the total train-exposure-hours for the full data set.)

A confidence interval is not easy to construct. Therefore, we leave frequentist estimation and proceed at once to Bayesian estimation.

6.7.1.2 Bayesian Estimation under Simple Assumptions

The simplest assumptions are that:

- the outages occur according to a Poisson process with frequency $\mathcal{O}_{freq}$, and
- the outage durations have an exponential($\mathcal{O}_{dur}$) distribution.

The first assumption is essentially true if the up-times have an exponential distribution and MTTF >> MTTR. Then the time from one outage onset to the next is dominated by the time up, which is exponentially

distributed. The parameter θ_{dur} can be interpreted as the reciprocal of the mean outage duration. Here, both parameters have units 1/time.

The conjugate prior for θ_{freq} can be denoted gamma($\alpha_{freq,0}$, $\beta_{freq,0}$), as stated in Section 6.2.2.4.1. The conjugate prior for θ_{dur} can be denoted gamma($\alpha_{dur,0}$, $\beta_{dur,0}$), as stated in Section 6.6.1.2.2. As above, suppose that n outages are observed in exposure time t_{expos}, with total outage duration t_{dur}. The posterior distributions are

$$\theta_{freq} \sim \text{gamma}(\alpha_{freq,1}, \beta_{freq,1})$$
$$\theta_{dur} \sim \text{gamma}(\alpha_{dur,1}, \beta_{dur,1}) ,$$

where

$$\alpha_{freq,1} = \alpha_{freq,0} + n$$
$$\beta_{freq,1} = \beta_{freq,0} + t_{expos}$$
$$\alpha_{dur,1} = \alpha_{dur,0} + n$$
$$\beta_{dur,1} = \beta_{dur,0} + t_{dur} .$$

By Equation 6.29 we have

$$q = \theta_{freq}/\theta_{dur} ,$$

which is proportional to the ratio of two chi-squared variables, because the chi-squared distribution is a re-expression of the gamma distribution. In addition, it is reasonable to think that the posterior distributions are independent; that is, if the outage frequency were to increase or decrease, this would provide no information about whether the mean outage duration increases or decreases. Therefore, q is proportional to the ratio of two *independent* chi-squared variables. However, the ratio of two independent chi-squared distributions, each divided by its degrees of freedom, has an F distribution, as is shown in many books on statistics and stated in Appendix A.7.11. It follows that q is distributed as

$$(\beta_{dur,1}/\beta_{freq,1})(\alpha_{freq,1}/\alpha_{dur,1})F(2\alpha_{freq,1}, 2\alpha_{dur,1}) .$$

The two quantities in parentheses following the F are the two parameters of the distribution. Selected percentiles of the F distribution are tabulated in many statistics books. They are not tabulated in this handbook because they are calculated by many software packages, including spreadsheets such as Microsoft Excel (2001) and Quattro Pro (2001). Facts about the F distribution are given in Appendix A.7.11.

If the Jeffreys noninformative priors are used, the posterior distribution of q is

$$(t_{dur}/t_{freq})[(n+0.5)/n]F(2n+1, 2n) . \qquad (6.30)$$

Table 6.17 shows 21 outages, with a total duration t_{dur} = 310.51 train-hours, and a total exposure time for the two trains t_{expos} = 17,850 train-hours. First, let us examine the assumptions, and then perform the calculations.

Section 2.6.2 lists an assumption about the independence of durations. Atwood and Engelhardt (2003) point out that Table 6.17 shows two instances when both trains were out for exactly the same amount of time in the same month. This is probably not coincidence, but indication that both trains were briefly taken out of service together, violating the independence assumption. However, the affected outage time is less than 1% of the total outage time, so Atwood and Engelhardt feel that the violation of the assumption is not serious.

Section 6.6.2.3.1 presents a Q-Q plot for checking whether durations follow an exponential distribution. The resulting plot for the present outage-duration data is given as Figure 6.61.

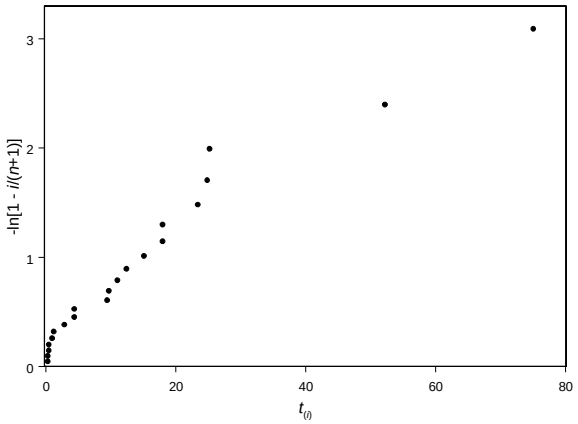


Figure 6.61 Q-Q plot for checking whether durations have exponential distribution in Example 2.16.

The line is not perfectly straight. The two largest times are a bit too large, and there are too many very small times. Thus, the true distribution is apparently more skewed than an exponential distribution. Nevertheless, we will assume that the exponential distribution is approximately correct, while recognizing that the resulting uncertainty intervals are not exact.

Atwood and Engelhardt use a similar plot to check the Poisson assumption for the outage count. More precisely, they investigate whether the times between outage onsets are exponential, which they should be if the outages follow a Poisson process. The outage onset times are not given in Table 6.17, but the authors assume that outages are approximately uniformly spaced in the month in which they

occur. Under this approximation, the times between outage onsets result in Figure 6.62.

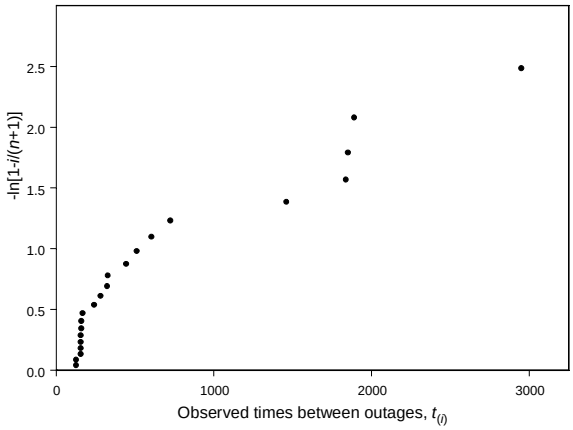


Figure 6.62 Q-Q plot for examining whether times between outages are exponential.

This figure shows the same general curvature as does Figure 6.61, but with even more curvature, suggesting that the distribution is more skewed than an exponential distribution. Atwood and Engelhardt recognize this fact but choose to press on with the analysis anyway.

Informative priors would be easy to use, but for this illustration we use the Jeffreys noninformative priors, resulting in a posterior distribution given by Expression 6.30. This says that the unavailability q has a distribution

(310.51/17850)(21.5/21) $F(43, 42)$.

The 5th and 95th percentiles of the F distribution with 43 and 42 degrees of freedom are 0.601 and 1.667. These percentiles are given by many software packages. Therefore, the 5th and 95th percentiles of q are 1.07E!2 and 2.97E!2. The mean is

(310.51/17850)(21.5/21) (42/40) = 1.87E!2,

using the formula for the mean of an F distribution from Appendix A.7.11. As a sanity check, this can be compared with the simple estimate given in Section 6.7.1.1, 310.51/17850 = 1.74E!2.

6.7.1.3 Model Validation

If the durations appear to be exponentially distributed, and the counts appear to have a Poisson distribution, one may then investigate the other assumptions listed in Section 2.6.2. Independence is difficult to verify from data, and is more likely verified by careful thinking. As for the assumption of a common distribution for all the down times and a common distribution for all the up times, the methods of Section 6.6.2 can be used: tools for investigating whether different portions of the data have different distributions, or whether a trend is present. The analyst must decide whether the data set is large enough to deserve this effort.

6.7.1.4 Bayesian Estimation under Other Assumptions

If the outages do not follow a Poisson distribution or if the durations do not follow an exponential distribution, Equation 6.29 can still be used, but it is much more difficult to estimate the two pieces of the equation. Atwood and Engelhardt (2003) discuss this issue very briefly. It will not be considered further here. Instead, we proceed now to methods that do not require assumptions on the forms of the distributions.

6.7.2 Analysis of Summary Data

The task now is to use the summary data only, not the data from individual outages, to obtain a Bayesian distribution for q. The fundamental technique is data aggregation to yield quantities X_i that are approximately independent and identically normally distributed, as described next. Atwood and Engelhardt (2003) are unable to prove theoretically that aggregation *must* result in normally distributed X_i. However, they perform simulations of representative cases and show that asymptotic normality occurs in those cases.

For ease of wording, we assume that the reporting periods are months, as they are in Example 2.16.

6.7.2.1 Data Aggregation

Denote the exposure time for the ith train-month by e_i and denote the corresponding outage time by o_i. The corresponding simple point estimate of the unavailability q is the ratio $x_i = o_i/e_i$. This gives one such estimate of q for each train-month of data. The estimate from any one train-month is not very good, because it is based on only a small data set. Indeed, if $e_i = 0$ the estimate is undefined.

The data may contain many zeros, as seen in Table 6.18. As a result of the many zeros and few relatively large outage times, the data can be quite skewed. To eliminate some of the zeros and make the data less skewed, the data can be pooled in various ways. For example, the rightmost column of Table 6.18 shows the total outage time for the two trains. Similarly, the data could be aggregated by time periods longer than one month, such as by calendar quarter or calendar year.

This aggregation over time could be done separately for each train or for the pooled data from the trains.

This aggregation groups the data into **subsets**, for example train-months (the least aggregation), or train-quarters, or system-months, etc. Note, we are still trying to estimate the train unavailability, not system unavailability, even if we pool the data from both trains in the system. Let o_i and e_i now denote the outage time and exposure time for the ith subset. The simple moment estimate of q based on the ith subset is $x_i = o_i/e_i$.

In the data of Table 6.18, if the two trains are pooled the total train exposure time for month 2 is e_2 = 720 + 720 = 1440 hrs, and the total train outage time is o_2 = 125.19 hrs. The estimate based on this one month is 125.19/1440 = 8.69E-2. If calendar quarters are pooled but trains are not pooled, the total train exposure time for Train 1 in quarter 3 is e_3 = 0 + 637 + 676 = 1313 hrs, and the corresponding train outage time is o_3 = 0 + 18.02 + 0 = 18.02 hrs. The estimate of q based on this one train-quarter is 18.02/1313 = 1.37E-2.

Whatever level of aggregation is used, this approach pools the numerators and denominators separately within each subset and then calculates the ratio.

The purpose of this aggregation is to produce multiple estimates o_i/e_i that we denote generically as x_i. The x_i values must all come from a single distribution. Therefore, the pooling assumes that the parameter q does not change within the data set, and that the various subsets of the data (calendar quarters or years, etc.) have similar exposure times, so that the random x_is all come from close to the same distribution.

In addition, the distribution of the x_is should be approximately normal. A normal distribution would not generate repeated values, such as multiple observed values of zero, nor would it produce strongly skewed data. Therefore, we must aggregate enough to obtain data that are not skewed and do not have repeated values.

How much aggregation is necessary? To investigate this question Atwood and Engelhardt (2003) perform some simulations, which indicate that minimal aggregation is not enough. In fact, the example data set should be many times larger than it is to make the method work really well. If an analyst has a large enough data set so that there is a choice between little aggregation into many subsets or much aggregation into few subsets, the second choice is the better one.

Table 6.19, from Atwood and Engelhardt (2003) gives some sample statistics for x, based on various amounts of aggregation of the data of Table 6.18. The skewness is a measure of asymmetry. Positive skewness corresponds to a long tail on the right. Zero skewness corresponds to a symmetrical distribution.

Table 6.19 Sample statistics for estimates of q, with different levels of aggregation.

	Train-month	System-month	Train-quarter	System-quarter
n	28	14	10	5
Mean	1.63E!2	1.63E!2	1.78E!2	1.78E!2
Median	0.00E+0	4.60E!3	1.38E!2	1.75E!2
St. dev., s	3.07E!2	2.51E!2	1.64E!2	1.16E!2
$s/n^{1/2}$	5.81E!3	6.70E!3	5.18E!3	5.19E!3
Skewness	2.79	2.02	1.25	0.33
No. zeros	17	7	2	0

The 28 values of x corresponding to train-months do not come from a normal distribution. They are too skewed, as is seen by the fact that the mean (1.63E-2) is very different from the median (0), and the skewness (2.79) is far from zero. Also, they have many occurrences of a single value, 0. Pooling the two trains into 14 subsets somewhat reduces the skewness and the percentage of zeros.

Pooling the three months for each train makes the distribution still more symmetrical: the mean and median are within 30% of each other, and the skewness is down to 1.25. When the data are aggregated by pooling trains *and* by pooling months into quarters, multiple values of zero are finally eliminated, and the distribution appears to be more nearly symmetrical: the mean and median are within 2% of each other, and the skewness is moderately small. This suggests that the five values of x_i may be treated as a random sample from a normal distribution.

To investigate this further, a Q-Q plot was constructed, as in Section 6.6.2.3.1. The plot given by Atwood and Engelhardt (2003) does not use the expected order statistics as in Figure 6.53, but instead is a different version of the plot. The authors found it easier to plot the ith order statistic against $M^{-1}[i/(n+1)]$, where M is the standard normal cumulative distribution function given in Table C.1. The plot is shown here as Figure 6.63.

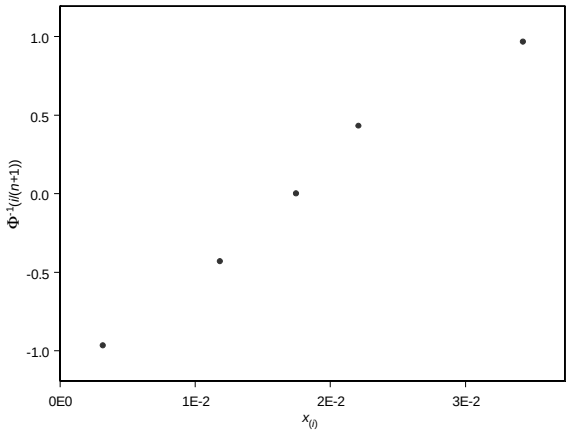


Figure 6.63 Q-Q plot for investigating normality of *x* when trains are pooled and months withing quarters are pooled.

The points lie almost on a straight line, indicating consistency with the normal distribution. For this reason, the authors treat the *x* values as coming from a normal distribution, when the data are aggregated to this degree. A goodness-of-fit test could be performed, but it would not be able to detect non-normality based on only five points.

As mentioned above, more aggregation would make the method work better. However, with the present very small data set, little more aggregation is possible.

There is a problem with the third quarter, because it has smaller exposure time than the other quarters. That means that *x* corresponding to this quarter has larger variance than the others. This is ignored here, but if the exposure time for quarter 3 had been even smaller, we might have dropped that quarter from the analysis, or pooled differently.

We repeat: The only purpose of the data aggregation is to eliminate the skewness and eliminate multiple values, thus permitting the use of normal methods. To the extent possible, we pool so that the aggregated subsets have similar exposure times, in order to have *x* values that come from a common distribution.

6.7.2.2 Frequentist Estimation

The same frequentist estimate of *q* can be used as with detailed data, the sum of the outage times divided by the sum of the exposure times. This ratio of the sums is not quite the same as the average of the ratios from the data subsets, because the various subsets are not based on the same exposure times. For example, in Example 2.16 quarter 3 has fewer exposure hours. Averaging the ratios for the subsets would treat the data from quarter 3 with as much weight as the data from the other quarters. Summing the outage times and exposure times first, before taking the ratio, gives quarter 3 only the weight that it should have.

A method to obtain a confidence interval for *q* uses facts about normally distributed random variables that are presented in many statistics books, and summarized here.

When $\{x_1, ..., x_n\}$ is a random sample from a normal(μ, σ^2) distribution, the usual estimates of μ and σ^2 are:

$$\bar{x} = \frac{1}{n}\sum_i x_i$$

and

$$s_X^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2 \quad .$$

(Note the $n - 1$ in the denominator, although some authors use n, and therefore use slightly different formulas below.)

The distribution of $\bar{X}$ is normal(μ, σ^2/n). When both μ and σ^2 are unknown, a $100(1-\alpha)\%$ confidence interval for μ is

$$\bar{x} \pm t_{1-\alpha/2}(n-1)s_X / \sqrt{n} \quad , \qquad (6.32)$$

where $t_{1-\alpha/2}(n-1)$ is the $1-\alpha/2$ quantile of the Student's *t* distribution with $n - 1$ degrees of freedom. For example, $t_{0.95}(n-1)$ gives a two-sided 90% confidence interval. Student's *t* distribution is tabulated in Appendix C, and is calculated by some software packages. Do not misread the $(n - 1)$ as a multiplier; it is a parameter, the degrees of freedom, of the Student's *t* distribution. In Table C.3 each row of the table corresponds to one value of the degrees of freedom.

In Example 2.16, we aggregate by system and calendar quarter, and use the resulting five values of *x* as if they are a random sample from a normal distribution. In the formulas above, the mean μ is the unavailability, *q*. From Expression 6.32 and Table 6.19, we obtain that a 90% confidence interval for *q* is

$$1.78E-2 \pm 2.132 \times 5.19E-3$$
$$= 1.78E-2 \pm 1.11E-2$$

because 2.132 is the 95th percentile of the Student's *t* distribution with 4 degrees of freedom. Thus, the

lower and upper confidence limits are

$q_{conf,0.05}$ = 7.E!3
$q_{conf,0.95}$ = 2.9E!2 .

This interval is approximate because the x values come from an approximately normal distribution.

6.7.2.3 Bayesian Estimation

Bayesian estimates are given here. Examples are worked out after the general formulas are given. As above, assume that the data have been aggregated enough so that $\{x_1, ..., x_n\}$ is a random sample from a normal($\therefore$, F^2) distribution.

6.7.2.3.1 Noninformative Prior

The joint noninformative prior for ($\therefore$, F^2) is proportional to $1/F^2$, as stated in Section 6.6.1.2.1 (in the context of lognormal distributions.) As stated in that section, the posterior distribution then results in

$$(\mu - \bar{x}) / (s_X / \sqrt{n})$$

having a Student's t distribution with n!1 degrees of freedom. Here $\therefore$ is the quantity with the Bayesian uncertainty distribution, and everything else in the expression is a calculated number. In the present context, $\therefore$ is the unavailability, q. It follows that the credible intervals agree exactly with the confidence intervals given by Expression 6.32. For example, a 90% credible interval for q is

$$\bar{x} \pm t_{0.95}(n-1)s_X / \sqrt{n} \ .$$

In Example 2.16, the mean $\therefore$ is interpreted as the unavailability q. Based on the values of x_i, the expression

(q ! 1.78E!2)/5.19E!3

has a Student's t distribution with 4 degrees of freedom. A 90% credible interval is

1.78E!2 ± 2.132×5.19E!3
= 1.78E!2 ± 1.11E!2
= (7.E!3, 2.9E!2)

which agrees exactly with the 90% confidence interval found above.

Not all PRA software packages contain Student's t distribution. Sometimes it is necessarily to temporarily adjust the method to match the available software. Analysts who are working without Student's t distribution in their software package may be forced to use a normal distribution with the same 90% interval as the one given by the above calculation. (To be more conservative, match the 95% intervals or the 99% intervals.) If the degrees of freedom are not too small ($ 3 as a bare minimum) the approximation of a Student's t by a normal is probably acceptable.

In the above example, a normal distribution with the same 90% interval would have 1.645F = 1.11E!2. Therefore, a normal approximation for the posterior distribution of q is normal with mean 1.78E!2 and standard deviation 6.75E!3.

6.7.2.3.2 Informative Priors

Informative conjugate priors for $\therefore$ and F^2 are presented in Section 6.6.1.2.1, along with the Bayesian update formulas. The prior parameters are:

d_0 = degrees of freedom for prior estimate of F^2
F_0^2 = prior estimate of F^2 (more precisely, $1/F_0^2$ is prior mean of $1/F^2$)
n_0 = prior number of observations (with variance F^2) for estimating $\therefore$
$\therefore_0$ = prior mean of $\therefore$

The update formulas are given in Section 6.6.1.2.1, resulting in four corresponding parameter values identified with subscript 1. The final result is that the posterior distribution of

$$(\mu - \mu_1) / (\sigma_1 / \sqrt{n_1})$$

is Student's t with d_1 degrees of freedom. Therefore, a 90% credible posterior interval is

$$\mu_1 \pm t_{0.95}(d_1)\sigma_1 / \sqrt{n_1} \ .$$

To develop an informative prior, Atwood and Engelhardt (2003) use some generic data from seven plants, with a mean unavailability for CVC trains of 6E!3 and a between-plant standard deviation 3.5E!3. Therefore, they set $\therefore_0$ = 6E!3 and F_0 = 3.5E!3. They point out that this data set does not perfectly match the (much older) data of Example 2.16. Therefore, they set n_0 = 3, not 7. They also argue that the between-plant variance has little or nothing to do with the between-calendar-quarter variance of Example 2.16. Therefore, they set d_0 = !1, corresponding to no prior information about the variance.

The update formulas of Section 6.6.1.2.1 yield:

$d_1 = -1 + 5 = 4$
$n_1 = 3 + 5 = 8$
$\therefore_1 = (3\times6E-3 + 5\times1.78E-2)/8 = 1.34E-2$
$F_1^2 = \{ -1\times(3.5E-3)^2 + (5 - 1)\times(1.16E-2)^2 +$
$[3\times8/(3+8)](6E-3 - 1.78E-2)^2 \}/4$
$= 2.07E-4$
$= (1.44E-2)^2$

Using the notation q instead of $\therefore$, it follows that

$(q - 1.34E-2)/[1.44E-2) / 2.83]$

has a posterior Student's t distribution with 4 degrees of freedom. A 90% posterior credible interval for unavailability is

$1.34E-2 \pm 2.132\times(1.44E-2)/2.83$
$= (3E-3, 2.4E-2)$.

6.7.2.4 Model Validation

A crucial feature of the simple method proposed above is aggregation of data, to reduce skewness and achieve approximate normality. The example analysis given above used a Q-Q plot, Figure 6.63, to check whether normality was achieved. In addition, other data plots and goodness-of-fit tests can be used to check the normality, as described in Section 6.6.2.3. However, if the data have been aggregated into a small number of sets, these tests and plots will probably not discover any departure from normality — there are too few data points. As mentioned above, to make the method work well, one should over-aggregate, beyond what the above analysis tools suggest is minimally acceptable.

An implicit assumption when pooling data subsets is that the various subsets correspond to the same distribution. Therefore, one may try to check this assumption, as follows.

The methods discussed in detail in Section 6.6.2.1 may be used, although the data may not be of good enough quality to show much. In particular, box plots may be used to suggest whether subsets can be pooled. The Kruskal-Wallis test, the nonparametric analogue of the analysis-of-variance test for equal means of normally distributed random variables, can be used to test equality of the unavailability in the data subsets. However, the many observations tied at zero make these methods difficult to apply.

6.7.3 Comparison of the Analyses with the Two Types of Data

The approach with detailed data works most easily if the number of outages follows a Poisson distribution and the outage durations can be assumed to have an exponential distribution. In that case, the posterior distribution of unavailability is a rescaled F distribution. The assumptions of Poisson counts and exponential outage durations must be checked.

The approach with summary data uses aggregation of the data into subsets until the estimated unavailabilities for the various subsets appear to be approximately normal. The resulting posterior distribution is a rescaled Student's t. Data aggregation eliminates the need to assume any particular distribution in the underlying process. However, one must confirm that approximate normality has been achieved.

Both methods work easily with either noninformative and informative priors, although the formulas for updating an informative prior with summary data are more intricate than when using detailed data.

The posterior distributions from the two analyses given above for Example 2.16 are compared in Table 6.20. It can be seen that in this example the two posterior distributions are quite consistent with each other.

Table 6.20 Comparison of Bayesian results from two approaches, using noninformative priors.

Data	Mean	90% Interval
Detailed	1.87E-2	(1.1E-2, 3.0E-2)
Summary	1.78E-2	(7.E-3, 2.9E-2)

7. TRENDS AND AGING

7.1 Overview

The material in this chapter is needed only if the model-validation tools in Section 6.2 or 6.3 have discovered the existence of a trend in an initiating-event rate θ or in a probability p. This chapter is more advanced than Sections 6.2 and 6.3, because it actually models the trend.

Such a trend might be in terms of calendar time, or in terms of system age. Section 7.2 considers trends in θ when the events are grouped into bins, such as counts of events in calendar years. Section 7.3 briefly considers trends in θ when the individual event times are used as the data. Section 7.4 considers trends in p. These sections all model θ or p as a parametric function of time. The final section of this chapter, Section 7.5, ties together some of the methods that have been presented in different specific applications in this chapter and in Chapter 6.

Modeling a trend normally involves some rather detailed mathematics. A Bayesian analysis must construct a sample from the posterior distribution, and a frequentist analysis must calculate the fitting equations and estimate the uncertainties in any unknown parameters. The viewpoint taken here is that the computer software will do those calculations. This chapter avoids any equations that the user will not need. Instead, this chapter presents the various approaches that are possible, and shows how to interpret the computer output and translate it, if necessary, into the quantities needed for probabilistic risk assessment (PRA).

It turns out that the Bayesian analysis is no harder to present than the frequentist analysis. Therefore, it is given first in Chapter 7.

Some of this material is drawn from an INEEL report by Atwood (1995), and indirectly from more advanced sources. The INEEL report gives additional examples, including definitions of somewhat more complex models and discussion of pitfalls in constructing such models.

7.2 Binned Poisson Data

7.2.1 Examples

Two examples are given here. The first example was introduced as Example 6.5, unplanned demands for the high pressure coolant injection (HPCI) system during 1987-1993. Table 6.8, summarizing that example, is repeated here for convenience as Table 7.1.

Table 7.1 HPCI demands and reactor-critical-years (from Example 6.5).

Calendar year	HPCI demands	Reactor-critical-years
1987	16	14.63
1988	10	14.15
1989	7	15.75
1990	13	17.77
1991	9	17.11
1992	6	17.19
1993	2	17.34

The second example, Example 7.1, groups events not by year of occurrence, but by age of the reactor.

Example 7.1 Thermal-fatigue leak events, by plant age.

Thermal-fatigue leaks in PWR stainless-steel primary-coolant-system piping are tabulated by Shah et al. (1998).

Age (years from initial criticality)	Number of leaks	Reactor-years
0.0 – 5.0	2	1052
5.0 – 10.0	1	982.5
10.0 – 15.0	4	756.9
15.0 – 20.0	4	442.4
20.0 – 25.0	2	230.9
25.0 – 30.0	0	43.9

The PWR plants considered here include all the western-designed PWRs, 217 reactors in all, from initial criticality until May 31, 1998, or until decommissioning. For details, see Shah et al. 1998. Plant age is summarized in 5-year bins, shown in the first column of the table. Other bins, such as 1-year bins or 10-year bins, could have been constructed. For each thermal-fatigue leak event, the age of the plant at the time of the event was calculated. The number of such events for each age bin is given in the second column of the table.

To count the reactor-years for each age, the number of reactors that experienced 1 year, 2 years, etc. were totaled. For example, Three Mile Island 2 had its initial criticality in December 1978, and was shut down in March 1979. It was counted as contributing 1/4 of a reactor-year (= 3 months) to the first bin (age 0.0 – 5.0). At the other extreme, Yankee Rowe operated from July 1961 to September 1991. It was counted as contributing 5 reactor-years in each of the six age bins. These counts of reactor-years are totaled in the third column of the table in Example 7.1.

The two examples have identical form. Each bin corresponds to a time, which can be coded numerically. This time will be called **clock time** here. In Example 6.5, the clock time is calendar time. Each bin corresponds to a year, and could run from 1987 through 1993, or from 87 to 93, or from 0 to 6, or even !3 to +3. Any coding works, as long as it is numerical. Denote the coded clock time for bin i by t_i. The units of t_i are calendar years, and the different possible codes correspond to different definitions of year 0. In Example 7.1, clock time is age. The bins are identified most simply by the midpoints of the age bins: 2.5, 7.5, 12.5, etc. The units of t_i are age in years. They could also be identified by other numerical codes, corresponding to the smallest age in the bin or the largest age in the bin or some other code.

In both examples, each bin has a clock time, an event count, and an exposure time. When considering possible trends, we must distinguish between clock time and exposure time, paying attention to a distinction that was not so important in the earlier chapters. In this chapter the symbols t_i, x_i, and s_i denote the coded clock time, the event count, and the exposure time, respectively, for bin i. Suppose that θ has units of events per reactor-year. Then the units of exposure time must be reactor-years. During any short time interval from t to $t + \Delta t$, the exposure time Δs equals Δt times the number of reactors operating during the time period.

To avoid redundancy, this section will use only the Example 6.5 for illustrating the methods, although either example could be used.

7.2.2 Model

7.2.2.1 General Model

The assumed model is an extension of the model for a Poisson process given in Section 2.2.2. The following assumptions are made. These are a simplification of the slightly weaker assumptions given by Thompson (1981):

1. The probability that an event will occur in any specified interval with short exposure time approaches zero as the exposure time approaches zero.
2. Exactly simultaneous events do not occur.
3. Occurrences of events in disjoint time periods are statistically independent.

This model is a **nonhomogeneous Poisson process (NHPP)**. The model in Section 2.2 is a **homogeneous Poisson process (HPP)**, a special case of the model given here. Consider now a single operating system, so that exposure time equals elapsed clock time. In the HPP, the probability of an event in the time interval $(t, t + \Delta t)$ is approximately $\theta \Delta t$. In the NHPP, θ is not constant, but is a function of t. The function $\theta(t)$ is called the time-dependent event occurrence rate. Some authors call it the **Poisson intensity function**. The probability of an event in the time interval $(t, t + \Delta t)$ is approximately $\theta(t) \Delta t$. In a longer interval, from a to b, the random number of events is Poisson with mean

$$\int_a^b \lambda(t)dt .$$

Four special cases are mentioned here.

1. The HPP has $\theta(t)$ = a constant > 0.
2. The **loglinear** model has $\ln\theta(t) = a + bt$, or equivalently, $\theta(t) = Ae^{bt}$, with $A = e^a$. This is also called the **exponential event rate** model. Here, a and b are unknown parameters, which must be estimated from the data.
3. The **Weibull process**, or **power-law event rate** model has $\theta(t) = (b/c)(t/c)^{b-1}$, or equivalently $\theta(t) = At^{b-1}$. Both b and c are unknown parameters (with $b > 0$, $c > 0$), to be estimated from the data. This model can be rewritten as

   $$\ln\theta(t) = \ln(A) + (b - 1)\ln(t) ,$$

   which is a linear function of $\ln(t)$. Several parameterizations are found in the literature.
4. An **extended power-law** process has $\theta(t) = At^{b-1} + \theta_0$, for $\theta_0 \geq 0$. The Swedish I-Book (Pörn et al. 1994) uses a Bayesian analysis of this model. The method is developed by Pörn (1990) and briefly explained by Pörn et al. (1993) and Lehtinen et al. (1997).

An occurrence rate must be nonnegative. Note that all four of these models have this requirement built in — they force $\theta(t)$ to be positive for all t. The number of possible models is infinite, because any nonnegative function of t gives a different model.

In the first bulleted case, $\delta(t)$ is constant, and in the other three cases $\delta(t)$ is monotone in t, either increasing forever or decreasing forever. An event frequency has a **bathtub** shape if $\delta(t)$ is decreasing when t is small, then roughly constant, and finally increasing when t is large. Models of a bathtub shape require more parameters. Therefore, bathtub curves are commonly used to describe models qualitatively, but have not been implemented in widely-used quantitative models.

As noted, the loglinear model satisfies

$$\ln\delta(t) = \ln(A) + bt$$

and the power-law model satisfies

$$\ln\delta(t) = \ln(A) + (b - 1)u$$

where $u = \ln(t)$. Therefore, the power-law model can be expressed as a loglinear model in $\ln(t)$, as long as t stays away from zero. Therefore, the illustrations of this section will use the loglinear model. If they so desire, readers can translate this material into the power-law model by redefining b and replacing t with $u - \ln(t)$. The extended power-law model will not be considered further here.

When multiple systems are observed simultaneously, the total number of events is again a Poisson random variable, and the mean count is the sum of the means for the individual systems. This fact will be used to account for the exposure times in the examples.

The loglinear and simple power-law models given above are now discussed in detail.

7.2.2.2 Loglinear Model

An occurrence rate must have units of events per exposure time unit. Therefore, in the loglinear model, A has units of events per exposure time unit, and b has the inverse units of the time axis. In Example 6.5, A has units 1/reactor-critical-year, and b has units 1/calendar-year. In Example 7.1, A has units 1/reactor-calendar-year, and b has units 1/year-of-age.

The loglinear model is illustrated here, for $a = 0.1$ and $b = +1.0$, 0.0, and -1.0. Figure 7.1 shows $\ln\delta(t)$ as a function of t, and Figure 7.2 shows $\delta(t)$ as a function of t.

The interpretation of b is the slope of $\ln\delta(t)$. If b is negative, $\ln\delta(t)$ is decreasing, and therefore the event occurrence rate $\delta(t)$ is also decreasing. If b is positive, $\delta(t)$ is increasing, and if $b = 0$, $\delta(t)$ is constant. Tests for trend will be formulated as tests of whether $b = 0$.

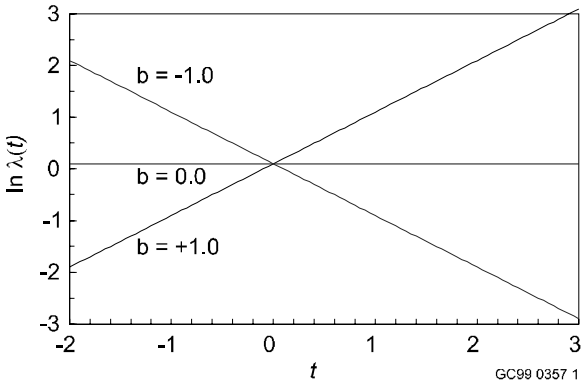


Figure 7.1 Loglinear model, $\ln\delta(t) = a + bt$, for $a = 0.1$ and three possible values of b. The vertical axis shows $\ln\delta(t)$.

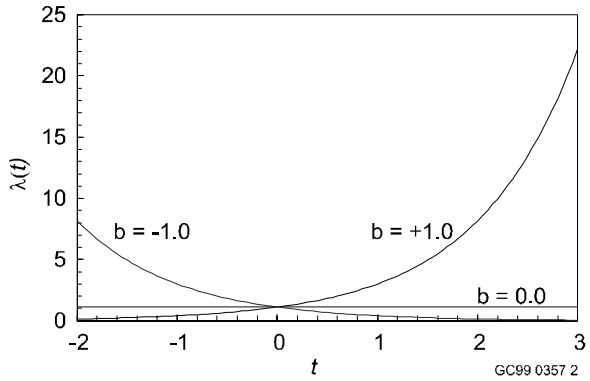


Figure 7.2 Same model as in previous figure, showing $\delta(t)$ instead of $\ln\delta(t)$.

The interpretation of a is the intercept of $\ln\delta(t)$, that is, the value of $\ln\delta(t)$ at $t = 0$. The meaning of a depends on how time is coded. In Example 6.5, with HPCI demands, if t runs from 1997 to 2003, the value $t = 0$ corresponds to about 2000 years ago, and a is the value of $\ln\delta(t)$ at that time. This is an enormous extrapolation, and a can be estimated only with great uncertainty. Coding t as running from 97 to 103 involves less extrapolation, because now $t = 0$ corresponds to the year 1900, only some 100 years before the data. Other possibilities are to let t run from 0 to 6, or from -3 to +3. These coding schemes involve no extrapolation at all, because 0 is included in the range of the observed data.

In theory, it makes no difference which coding system is used. The different codings for t and the different meanings of a compensate for each other. For any particular time, such as 1996 or 2001, the different coding systems give exactly the same estimate of δ at that time. In practice, however, no computed value is exact, and roundoff errors can accumulate. Use of large

extrapolations can introduce large errors. Well-programmed software should protect against this problem automatically, no matter how the times are entered. Nevertheless, the analyst can do no harm by choosing a coding system with zero reasonably close to the data.

7.2.2.3 Power-Law Model

In both parameterizations given above for the power-law model, b is a unitless shape parameter. As will be seen below, $b - 1$ is the slope of $\ln\lambda(t)$ as a function of $\ln(t)$. In the first parameterization, c is a scale parameter with units of t. It does not have a simple interpretation. In the second parameterization, A has strange units, but a simple interpretation as the numerical value of $\lambda(t)$ at $t = 1$.

Figures 7.3 and 7.4 illustrate the power-law model for $A = 1.0$, and $b = 0.5$, 1.0, 2.0, and 3.0.

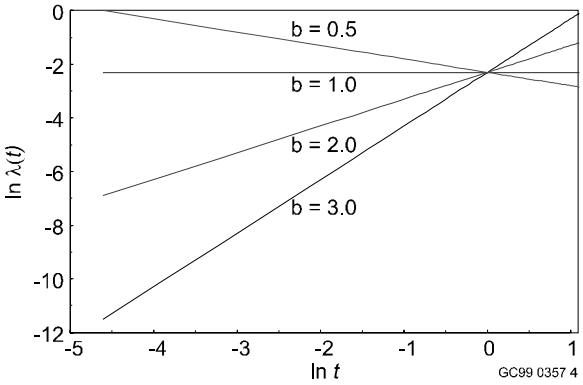


Figure 7.3 Power-law model, showing $\ln\lambda(t)$ as a linear function of $\ln(t)$, with $A = 1.0$ and several values of b.

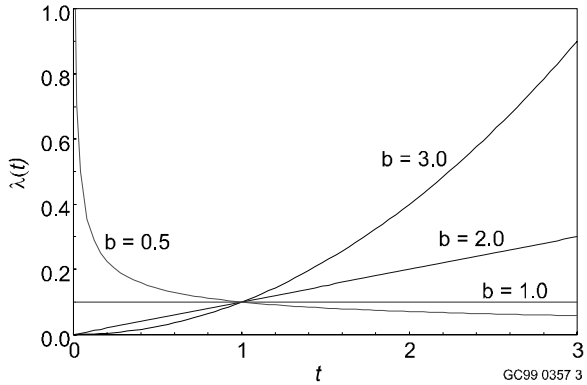


Figure 7.4 Same model as in previous figure, with $\lambda(t)$ shown as function of t.

In Figure 7.3, the parameter $b - 1$ is the slope of $\ln\lambda(t)$ as a function of $\ln(t)$. In Figure 7.4, b is a shape parameter, defining the shape of the curve. In either figure, the interpretation of A is the numerical value of $\lambda(t)$ at $t = 1$.

This model requires $t \geq 0$, so a coding system should be chosen so that all the observed times correspond to nonnegative values of t. Also, if the occurrence rate is decreasing, the modeled occurrence rate becomes infinite at $t = 0$.

The loglinear and power-law models are widely used, but they are chosen for their simplicity and convenience, not their theoretical validity. Any model must be checked for goodness of fit. Moreover, no model should be extrapolated far into the future — even if some convenient algebraic formula fits a trend well in the past, that is no guarantee that the data will continue to follow that formula in the future.

7.2.3 Bayesian Estimation with Loglinear Model

The first few paragraphs here describe the big picture in very general terms. Following that, the section carries out the Bayesian estimation when the occurrence rate satisfies the equation $\ln\lambda(t) = a + bt$.

A large-sample approximation is applicable. As the observed event counts become large, the form of the likelihood function approaches the form of a normal density in the unknown parameters. That is, if the likelihood function were treated as a probability density, it would be approximately a normal density. This is a general fact for large data sets that is exploited by advanced statistics texts, such as Cox and Hinkley (1974, Section 10.6). Therefore, with large data sets the **conjugate** prior is normal: if the unknown parameters are given a normal prior distribution, the posterior will be approximately normal, with very good approximation as the data set becomes large. The corresponding **noninformative** prior for a and b is the limiting case as the variance approaches infinity, which is a constant density.

For the work here, it will be assumed that a computer program produces a sample from the posterior distribution. The theory sketched above then leads us to the conclusion that the posterior distributions not only *appear* normal, they really *are* normal, or very close to normal.

Now let us move to the specific case at hand, with $\ln\lambda(t) = a + bt$, and with a and b as the unknown parameters. For this case it happens that the above normal approximation is valid when the event counts are only moderate in size.

The bins must be small enough that $\lambda(t)$ is approximately a straight line function within each bin, not strongly curved within the bin. Denote the midpoint of the ith bin by t_i. Then the expected number of events in the bin is well approximated by $\lambda(t_i)s_i$, where s_i is the exposure time for the bin. The method is to fit the observed Poisson counts to $\lambda(t_i)s_i$, while assuming that $\lambda(t_i)$ has the form $a + bt_i$.

A convenient software package is BUGS (1995), Bayesian inference Using Gibbs Sampling. The Windows version is called WinBUGS. It is also described in Section 8.2.3, and documented by Spiegelhalter et al. (1995). It is currently available for free download at

http://www.mrc-bsu.cam.ac.uk/bugs/ .

WinBUGS is a high-powered research tool, capable of analyzing very complex models. It does this by not trying to obtain a simple random sample from the posterior distribution. Instead, it tries for something more restricted, a **Markov chain Monte Carlo (MCMC)** model. Here a **chain**, or sequence, of numbers is generated, starting at an arbitrary point but eventually sampling from the posterior distribution. The values in the sequence are not independent, but this does not matter. After the initial value has been essentially forgotten, the remaining values form a sample from the posterior distribution. They can be used to approximate the moments, percentiles, and shape of the distribution.

WinBUGS can be used either with a graphical description of the model, called a "directed graph," or with a text script. Example 6.5 will be analyzed using WinBUGS here, assuming a loglinear model.

With the data of Example 6.5 (Table 7.1), BUGS was used to model $\lambda(i) = \exp(a + bi)$, for i from 1 to 7. Then $X(i)$ was modeled as having a Poisson distribution with mean $\mu(i) = \lambda(i)\times s(i)$. Finally, a and b were given very diffuse prior normal distributions. Figure 7.5 is the logical diagram showing the relations.

In Figure 7.5, deterministic relations are shown by dashed arrows, and stochastic relations (random number generation) are shown by solid arrows. The parts of the model that depend on i are enclosed in the box.

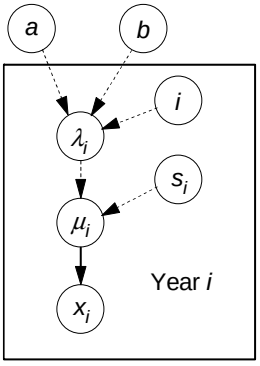


Figure 7.5 Directed graph for analysis of Poisson trend in Example 6.5.

Figure 7.6 shows the BUGS script that was used. Many users find the text script easier to manipulate than the graph.

One idiosyncrasy of BUGS is that it parameterizes the normal distribution in terms of the precision $J = 1/F^2$. The reasons are explained in Section 6.6.1.2.1. Therefore, a precision of 0.0001 in the script corresponds to a standard deviation of 100. That gives a very diffuse distribution.

```
model
{
   for (i in 1:N) {
      lambda[i] <- exp(a + i*b)
      mu[i] <- lambda[i]*s[i]
      x[i] ~ dpois(mu[i])
   }
   a ~ dnorm(0.0, 0.0001)
   b ~ dnorm(0.0, 0.0001)
}
```

Figure 7.6 BUGS script for analyzing data of Example 6.6.

The script was executed with four separate starting values of a and b, generating four chains of values, each 10,000 elements long. The first 200 elements of each chain were discarded, and the remaining 39,200 elements were used to estimate the posterior distributions. Table 7.2 summarizes the posterior distributions for a and b. When interpreting these summaries, be aware that a and b are not independently distributed.

Even though the numbers are not necessarily accurate to three places, the table shows that the mean and median are nearly equal in each case, and the 5th and 95th percentiles are approximately 1.645

standard deviations from the mean. That is, the distributions appear to be approximately normal. (BUGS supplies graphical estimates of the densities, which also appear normal.) This is consistent with the theory mentioned earlier.

Table 7.2 Posterior statistics for *a* and *b*, for loglinear model of Example 6.5.

	a	*b*
mean	0.264	!0.237
median	0.27	!0.237
st. dev.	0.251	0.067
5th percentile	!0.157	!0.348
95th percentile	0.662	!0.129

The mean of *b* is negative, and 3.5 standard deviations away from the mean. This is very strong evidence of a downward trend. The posterior belief in a flat or rising trend is only 2.3E!4. (This is M(!3.5), from Table C.1.)

The posterior distribution of θ is lognormal in any year. It is shown in Figure 7.7. The median is plotted as a solid line and the 5th and 95th percentiles are shown as dashed lines. The simple point estimates

$$\hat{\lambda} = x_i / s_i$$

are plotted as dots.

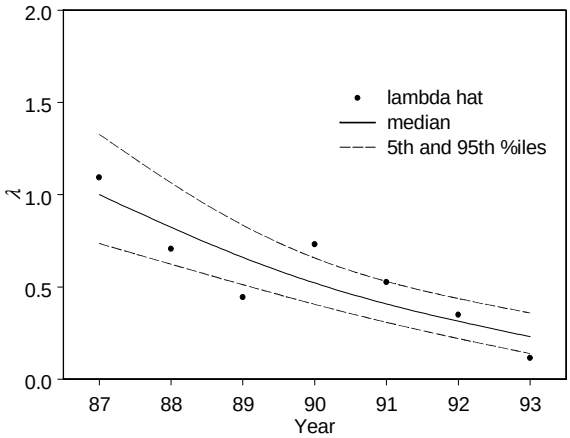


Figure 7.7 Posterior distribution of θ, assuming exponential trend in Example 6.5.

For *any particular year*, the value of $\theta(t) = \exp(a + bt)$ is between the two dotted lines with 90% probability. This is enough for many applications.

Suppose, however, that we were interested in the *entire curve*. This curve is the set of two-dimensional points of the form

$$\{ (t, \theta(t)) \mid !4 < t < 4 \} .$$

For two distinct times t_1 and t_2, a pair (a, b) that puts $\theta(t_1)$ between the lines may put $\theta(t_2)$ outside the dotted lines. Therefore, the entire curve does not fall between the two dotted lines with 90% probability. A 90% region for the entire curve would need to be wider than the band shown in Figure 7.7. This subtle issue is revisited for frequentist estimation in Section 7.2.4.5.

7.2.4 Frequentist Estimation with Loglinear Model

The frequentist method has several variations, which have been implemented in various software packages. They are presented here, applied to the example of Table 7.1, and the results are compared to each other and to the Bayesian results.

Assume a loglinear model, $\ln\theta(t) = a + bt$. Statistical software packages present their products using some technical terms, summarized here.

- **General linear model**: the mean of the observable random variable is a linear function of unknown parameters. This is NOT useful for the present problem. It is mentioned only to point out the possible confusion with the generalized linear model below.
- **Loglinear model**: the logarithm of the mean of the observable random variable is a linear function of unknown parameters. This is exactly the model considered in this section.
- **Generalized linear model**: a transformation of the mean of the observable random variable is a linear function of unknown parameters. This includes the loglinear model as a special case, when the transformation is chosen to be the logarithm.

7.2.4.1 Point Estimation

Analysis of the loglinear model finds the maximum likelihood estimates (MLEs) of *a* and *b*, based on the Poisson counts. The discussion below will sometimes call this the **Poisson-maximum-likelihood** method.

This approach is applied here to Example 6.5. No calculations are given for Example 7.1 because they would be very similar, differing only by using age instead of calendar year.

A loglinear model was fitted to the HPCI demand data of Table 7.1 (Example 6.5). When the years were coded as 87 through 93, the estimates of a and b were 20.5305 and !0.2355. The second number is the slope of $\ln\delta(t)$, and the first is the fitted value of $\ln\delta(t)$ when t = 0, that is, in the year 1900. (Of course, no HPCI systems existed in 1900, but the model does not know that and fits a value.) When, instead, the years are coded from 0 to 6, the slope is the same, but the intercept parameter is different, because now the value t = 0 is the year 1987. The estimate of a, the intercept, is 0.0389, the fitted value of $\ln\delta(t)$ for 1987.

The fitted value of $\delta(t)$ is the same, whichever coding method is used. The fitted values are shown in Figure 7.8, overlaid on Figure 6.22. Each point and vertical confidence interval is based on data from a single year, but the fitted trend uses all the data.

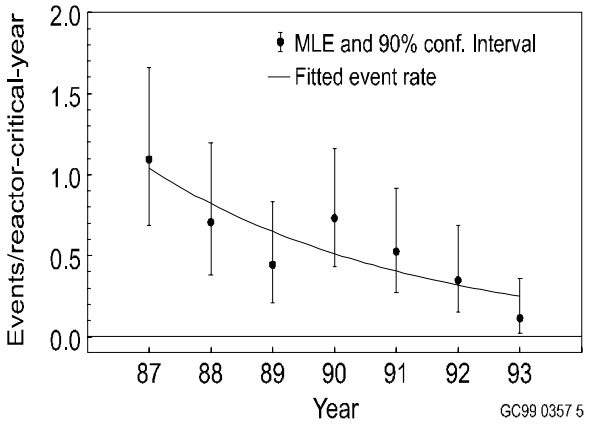


Figure 7.8 Frequency of unplanned HPCI demands, from Figure 6.22, with exponentially decreasing fitted trend line overlaid.

7.2.4.2 Confidence Intervals for *a* and *b*

With the point estimates $\hat{a}$ and $\hat{b}$, almost all software will also report standard errors, estimates of the standard deviations of the estimators. The estimators are assumed to be approximately normally distributed. This is a good approximation if the number of observations in each bin is not too small. (One rule of thumb is that the count in most bins be at least five. This is based on the observation that each bin's Poisson distribution is approximately normal if the mean is five or more. This rule of thumb is sufficient, but perhaps unnecessarily conservative.)

A 100(1 ! α)% confidence interval for b is

$$\hat{b} \pm z_{1-\alpha/2}\,se(b) \qquad (7.1)$$

where $\hat{b}$ is the estimate, and $z_{1-\alpha/2}$ is the 1 ! α/2 quantile of the normal distribution. For example, for a 90% confidence interval, α equals 0.1 and the 0.95 quantile is 1.645. The term $se(b)$ is the corresponding standard error of b, the estimated standard deviation of $\hat{b}$.

A confidence interval for a is constructed in a similar way, but is normally much less interesting. Who cares what value a has? That parameter is just the intercept at some arbitrarily coded time with t = 0. The parameter a is of interest only because it can be used to construct confidence intervals for $\ln\delta(t) = a + bt$.

7.2.4.3 Test for Presence of Trend

Let two hypotheses be defined by:

H_0: $\delta(t)$ is constant.
H_1: $\delta(t) = \exp(a + bt)$, $b \neq 0$.

The loglinear model is used as an illustration for the alternative hypothesis, but any other specific model could be used, as long as it is not constant.

Note that $\delta(t) = \exp(a + bt)$ is constant if and only if b is zero. Therefore, the test of H_0 is the same as a test that b = 0.

As mentioned in Appendix B, tests for hypotheses about b are intimately related to confidence intervals for b. The hypothesis

H_0: $b = b_0$

is rejected in favor of the hypothesis

H_1: $b \neq b_0$

at significance level α if and only if the 100(1 ! α)% confidence interval does not contain b_0. In particular, the hypothesis

H_0: b = 0 ,

the hypothesis of no trend, is rejected at level 0.10 if the 90% confidence interval for b is entirely on one side of 0. The hypothesis is rejected at level 0.05 if the 95% confidence interval is entirely on one side of 0, and so

forth. Most software packages print out a significance level at which H_0: $b = 0$ is rejected, the p-value for the trend.

Similarly, software packages typically print a significance level at which the hypothesis $a = 0$ is rejected. This should be ignored, because the value of a has no inherent interest.

We now compare the above test to an earlier one. Section 6.2.3.2.2 gave tests for the presence of a trend, illustrated with Example 6.5. The only test it gave with binned data was the chi-squared test of:

H_0: $\delta(t)$ is constant.
H_1: $\delta(t)$ is not constant.

Section 6.2.3.2.2 commented that the test is not very powerful, because it considers such a broad class of possibilities as the alternative hypothesis.

In Example 6.5, the chi-squared test rejected the hypothesis of constant δ with p-value 0.009. The present test of $b = 0$ rejects this hypothesis with p-value 0.0004. Although both tests reject H_0, the test based on the loglinear model finds stronger evidence against constant δ than the chi-squared test did. In an example with a less clear trend, the test based on $b = 0$ might find a statistically significant trend when the chi-squared test did not.

Suppose that δ were not constant, but went up and down in an irregular way with no persistent increasing or decreasing trend. The chi-squared test might discover this, but the test based on b would not discover the nonconstancy of δ — more precisely, the test of $b = 0$ might "discover" the nonconstancy because the random data might appear to indicate a trend in spite of the true non-trending pattern of δ, but this would only be an accident. The test to use depends on the alternatives that the analyst regards as credible. A test that focuses on those alternatives will be more powerful than a test that is designed for different alternatives.

7.2.4.4 Confidence Interval for $\delta(t)$ at Fixed t

Most software packages also can find approximate confidence intervals for $\ln\delta(t)$ at particular values of t. It is worthwhile understanding the approach, because the software output may require modification to coincide with the analyst's needs. The idea is that the MLEs $\hat{a}$ and $\hat{b}$ are approximately normally distributed. The software finds an approximate $100(1 - \alpha)\%$ confidence interval for $\ln\delta(t)$ as

$$\hat{a} + \hat{b}t \pm [z_{1-\alpha/2} \times se(a + bt)] \qquad (7.2)$$

where $z_{1 - \alpha/2}$ is as defined earlier, and $se(a + bt)$ is the standard error, the estimated standard deviation of $\hat{a} + \hat{b}t$. The standard error depends on the value of t. It is found by the software — it cannot be found in a naive way from the standard errors of a and b, because the MLEs $\hat{a}$ and $\hat{b}$ are correlated, not independent. Expression 7.2 is a confidence interval for $\ln\delta(t)$. The confidence interval for $\delta(t)$ itself is found by taking the exponential of the two bounds in Expression 7.2.

Understanding this algebraic form may be useful. For example, suppose that the software insists on giving only a 95% confidence interval for $\delta(t)$, and the analyst desires a 90% interval instead. The following modification can be made. The α corresponding to a 95% confidence interval is 0.05. First, take logarithms of the reported upper and lower confidence limits for $\delta(t)$. Use these two values, and the form of Expression 7.2, to find

$z_{0.975} \times se(a + bt)$.

This follows from the fact that a 95% confidence interval corresponds to $1 - \alpha/2 = 0.975$. Using $z_{0.975}$ and $z_{0.95}$ from a table of the normal distribution, calculate the value of

$z_{0.95} \times se(a + bt)$.

From this, calculate the 90% confidence interval for $\delta(t)$,

$$\exp[\hat{a} + \hat{b}t \pm z_{0.95} se(a + bt)] \ . \qquad (7.3)$$

7.2.4.5 Simultaneous Confidence Band at All t

The above confidence interval is for a particular t. For many values of t, many such confidence intervals could be calculated. Each is a valid confidence interval, but they are not simultaneously valid. This is a subtle point. To appreciate it, recall the interpretation of the 90% confidence interval for $\delta(t)$ for some particular time t_1:

Pr[confidence interval for $\delta(t_1)$ contains true occurrence rate at time t_1] = 0.90. (7.4)

Here, the data set is thought of as random. If many data sets could be generated from the same set of years, each data set would allow the calculation of a confidence interval for $\delta(t_1)$, and 90% of these confidence intervals would contain the true occurrence rate.

A similar confidence statement applies to each time. The simultaneous statement would involve

Pr[confidence interval for $\delta(t_1)$ contains true occurrence rate at time t_1 AND
confidence interval for $\delta(t_2)$ contains true occurrence rate at time t_2 AND
so forth]. (7.5)

This probability is hard to quantify, because the intervals are all calculated from the same data set, and thus are correlated. However, Expression 7.5 is certainly smaller than 0.90, because the event in square brackets in Expression 7.5 is more restrictive that the event in brackets in Equation 7.4.

This problem is familiar in the context of least squares fitting. For example, Neter and Wasserman (1974) discuss it, and attribute the solution to Working, Hotelling, and Scheffé. A simple adaptation to the present setting is sketched by Atwood (1995, App. B-7). The simultaneous confidence band is obtained by replacing $z_{0.95}$ in Expression 7.3 by

$$[\chi^2_{0.90}(r)]^{1/2},$$

where the expression in square brackets is the 90th percentile of a chi-squared distribution with r degrees of freedom. Here, r is the number of unknown parameters, 2 in this example.

The simultaneous 90% confidence band is about 30% wider than the band of 90% confidence intervals, because

$[\chi^2_{0.90}(r)]^{1/2} = 2.15,$

which is about 30% larger than

$z_{0.95} = 1.645$

when $r = 2$.

Figure 7.9 again uses the HPCI unplanned-demand data of Example 6.5. The annual estimated event frequencies are shown, along with the fitted frequency, the simultaneous 90% confidence band on the frequency, and the band constructed from the individual 90% confidence intervals.

Simultaneous confidence bands typically are not calculated by software packages. They can be calculated by the user, however, based on the formulas above and the information produced by the software package.

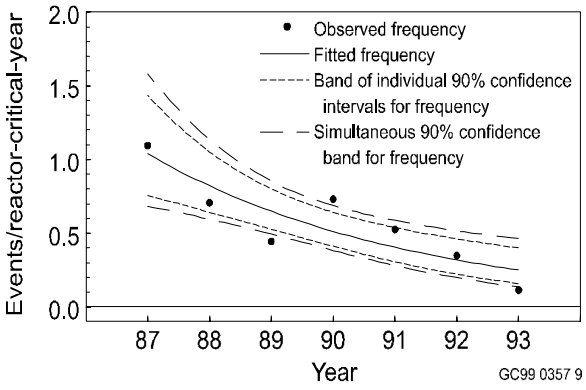


Figure 7.9 Simultaneous 90% confidence band and band formed by individual 90% confidence intervals for Poisson event rate, $\delta(t)$.

Which should be presented to users, the simultaneous band or the band of individual confidence intervals? If the user's interest is in a single time, such as the most recent time, then clearly the confidence interval at that time is of greatest interest. If, on the other hand, the user will look at the entire plot, for example to judge the existence of a trend, then the simultaneous confidence band is a better indication of the uncertainty in the estimated line. To satisfy both types of users, the graph could show the simultaneous confidence band and the confidence interval at the time of greatest interest. Figure 7.10 shows such a plot, assuming that a user would be most interested in the most recent time, 1993.

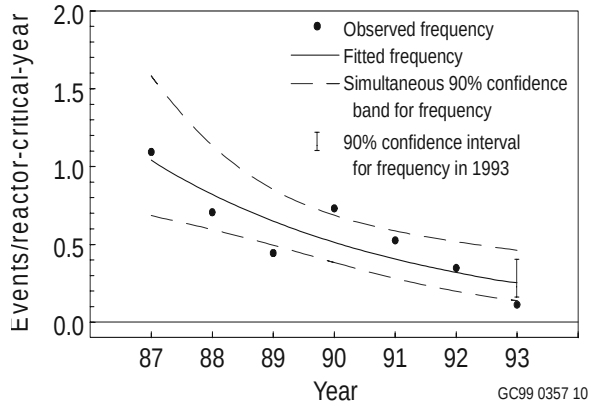


Figure 7.10 Simultaneous 90% confidence band for $\delta(t)$, and 90% confidence interval for one frequency of special interest, $\delta(93)$.

7.2.4.6 Alternative Using Least-Squares

Since the model assumes

$\ln\delta(t) = a + bt$,

one might decide simply to use least-squares software as follows. First, estimate λ based for each bin, based on only the data for that bin;

$$\hat{\lambda}_i = x_i / s_i$$

Then fit $\ln \hat{\lambda}_i$ to $a + bt_i$ by least squares. In principle, this works. In practice, the method has several twists in the road, described next.

First, if the observed count is zero in any bin, the MLE $\hat{\lambda}$ will be zero for that bin, and the logarithm will be undefined. This is the case for the final bin of Example 7.1. The following ways around this have been proposed.

- Instead of estimating λ_i by x_i/s_i, use $(x_i + ½)/s_i$. This is equivalent to replacing the MLE by the posterior mean based on the Jeffreys noninformative prior.
- Estimate λ_i by the posterior mean based on the constrained noninformative prior. In this case, the constraint could be that the prior mean equals the observed overall mean $\Sigma_i x_i/\Sigma_i s_i$, or equals a modification to guarantee a positive number, $(\Sigma_i x_i + ½)/\Sigma_i s_i$.

Such ways tend to reduce the trend slightly, because they add a constant to all the failure counts, slightly flattening out any original differences.

The second point that must be considered is that the variance of X_i/s_i is not constant. Ordinary least-squares fitting has some optimality properties if the variance is constant. Otherwise, it is more efficient to use weighted least squares, with weights inversely proportional to the variances of the observations. Many statistical software packages perform weighted least squares estimation.

For simplicity, this issue will be explained for the case with no zero counts, and with $\ln\lambda_i$ estimated by $\ln(\hat{\lambda}_i) = \ln(x_i/s_i)$. The variance of $\ln(X_i/s_i)$ is approximately the **relative variance** of X_i/s_i, defined as $\text{var}(X_i/s_i)/E^2(X_i/s_i)$. This is $1/E(X_i) = 1/(\lambda_i s_i)$ if X_i has a Poisson$(\lambda_i s_i)$ distribution.

Unfortunately, the variances depend on the λ_i values, which are unknown. Therefore, the following **iteratively reweighted least-squares** method can be used. Begin by assuming that λ is constant, and fit $\ln(x_i/s_i)$ to a straight line with weighted least squares, and weights s_i. Calculate the resulting estimates of λ_i, $\hat{\lambda}_i = \exp(\hat{a} + \hat{b}t_i)$. Then refit the data to a straight line, using weighted least squares, and weights $\hat{\lambda}_i s_i$. Repeat this process until the estimates stabilize.

The final point is that least-squares fitting typically assumes that the data are approximately normally distributed around the straight line. In the present context, this means that $\ln(X_i/s_i)$ is assumed to be approximately normally distributed. This assumption is acceptable, unless the mean count is close to zero. The variance of the normal distribution is then estimated from the scatter around the fitted line. This differs from the typical treatment of Poisson data, where the mean determines the variance.

A 90% confidence interval for $\lambda(t)$ at a particular t is given by

$$\exp\{\hat{a} + \hat{b}t \pm [t_{0.95}(d) \times se(a + bt)]\} \qquad (7.6)$$

where $t_{0.95}(d)$ is the 95th percentile of Student's t distribution with d degrees of freedom. The software will report the value of d. It is the number of bins minus the number of estimated parameters, 7 ! 2 in Example 6.5. The form of this equation is very similar to the form of Equation 7.3, although the estimates and standard deviation are calculated somewhat differently.

A simultaneous 90% confidence band has the same form, but the multiplier $t_{0.95}(d)$ is replaced by

$$[2F_{0.90}(r, d)]^{½},$$

where $F_{0.90}(r, d)$ is the 90th percentile of the F distribution with r and d degrees of freedom. This modification of Equation 7.6 is analogous to that for Equation 7.3 to get a simultaneous confidence band.

7.2.5 Comparison of Methods

The three frequentist methods are compared here. Following that comparison, the Bayesian method is compared to the frequentist methods, first for the current example and then in general.

Figure 7.11, from Atwood (1995), shows results from three frequentist analyses of the data of Table 7.1. As can be seen in the figure, the fitted lines are similar for all three analyses. The unweighted-least-squares method gives an unnecessarily wide confidence band. This shows the inefficiency of unweighted least squares when the variance is not constant. The Poisson-maximum-likelihood ap-

proach has a slightly narrower confidence band than does the weighted-least-squares approach. The reason is that the least-squares approach introduces an additional parameter that must be estimated, the variance around the line. The pure Poisson approach calculates the variance from the fitted mean. The price of estimating an extra parameter is a larger multiplier, $[2F_{0.90}(2, 5)]^{1/2} = 2.75$ instead of $[\chi^2_{0.90}(2)]^{1/2} = 2.15$.

One other reason for the difference between the two plots on the right of the figure might be that they use different estimators of the variance. In this example, however, the scatter around the line agrees almost perfectly with the scatter predicted by the Poisson assumptions, so the two calculations of the variance agree. This is coincidence, but it eliminates a possible distraction in comparing the two methods.

In summary, all three calculations are valid, and yield similar fitted lines. The method of unweighted least squares uses the data in a somewhat inefficient way, however, and therefore produces an unrealistically wide confidence band. The method of iteratively reweighted least squares provides a narrower confidence band, and the loglinear model provides the narrowest band of all. None of the calculations is exact: the least-squares method treats ln(count/time) as approximately normal, and the Poisson-maximum-likelihood method treats the parameter estimators as approximately normal. However, the analysis based on the Poisson-maximum-likelihood method is preferable (if the Poisson assumption is correct), because it gives the tightest confidence band, and reweighted least squares is second best. It may be that extra sources of variation, or "overdispersion," have entered the data, variation beyond that caused by the Poisson distribution. If so, reweighted least squares would be best and the Poisson-maximum-likelihood method would produce an unrealistically narrow band.

Now the frequentist methods are compared with the Bayesian method, first for the particular example, and then in general.

The corresponding figure from the Bayesian analysis is Figure 7.7. Careful examination of the figures shows the following:

- The posterior median in Figure 7.7 is close to the fitted line (the MLE) in Figure 7.9, which is the middle panel of Figure 7.11.
- The 90% credible band in Figure 7.7 shows a band that is valid for any one time, but not simultaneously for all time. It is close to the inner band in Figure 7.9, which also is valid for any one time but not simultaneously. These bands are somewhat narrower than the simultaneous bands of Figure 7.11.

The following comments apply in general, not just to the example:

- Frequentist estimation relies on approximate normality of the estimators, and therefore does not work well with data having few observed events. Bayesian estimation obtains normal posteriors

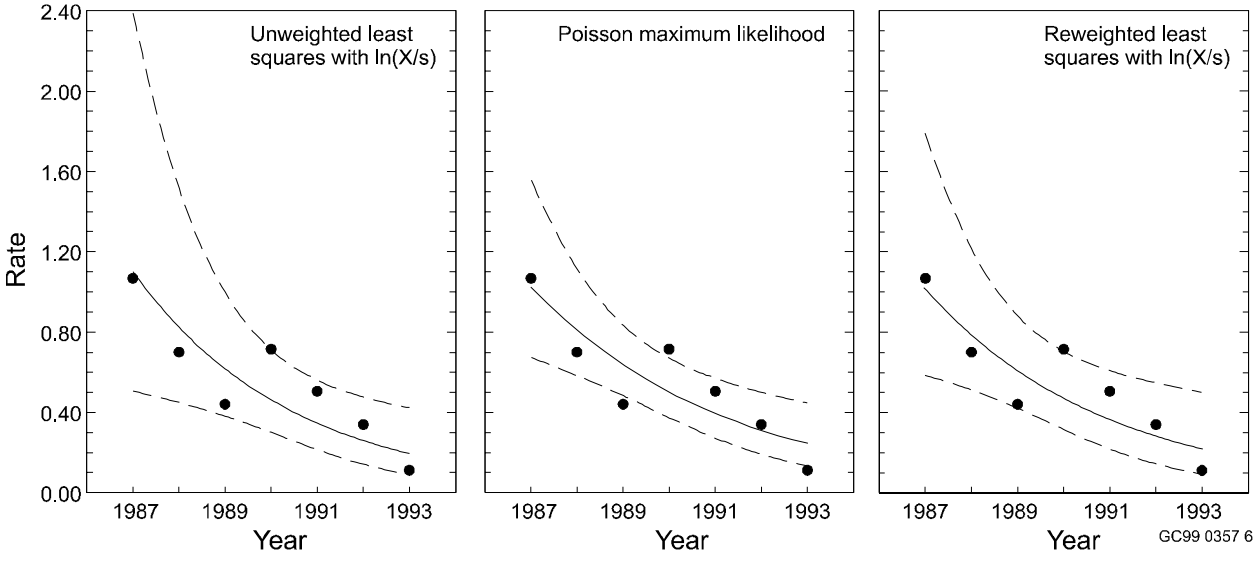


Figure 7.11 Fitted Poisson event occurrence rate and simultaneous 90% confidence band, based on three ways of fitting the HPCI unplanned-demands data.

when the data set is large, but does not fail entirely when the data set is small – it merely obtains a different, non-normal, posterior.

- Most frequentist software packages for analyzing trends include calculations for investigating the goodness of fit of the model, as will be seen in Section 7.2.6. Current Bayesian software may neglect this issue of model validation.

7.2.6 Model Validation

The three assumptions for a nonhomogeneous Poisson process are given at the beginning of Section 7.2.2. The first assumption is difficult to test from data. The second, dealing with common cause failures, has been addressed in Sections 2.2.4 and 6.2.3.3. The third assumption is that event counts in disjoint intervals are independent. This was addressed using a test of serial correlation in Section 6.2.3.4 when β is constant, but the analogue is too complicated to consider here. When the data are collected into bins, as assumed here, serial dependence can result in an unusually high or low event count in a single bin. This can be discovered by goodness-of-fit tests, considered below.

The final assumption made when fitting a trend model is the form of the model. Methods to examine the goodness of fit will be illustrated with the loglinear model, following the precedent of the rest of this section.

7.2.6.1 Graphical Check for Goodness of Fit

The natural graphical check is to compare the observed values to the fitted line. This is illustrated by Figure 7.8. In that figure, the 90% confidence interval for each year overlaps the fitted trend line. Because no year deviates strongly from the overall trend, the data appear consistent with the assumption of an exponential trend. Even if one 90% interval had failed to overlap the fitted trend line, one would not necessarily conclude that the exponential-trend assumption is violated. The reason for not being concerned about a single failure to overlap is that some 90% confidence intervals are expected to miss the true value. In the long run, as many as 10% of the intervals may fail to contain the true value. In the short run, one miss in seven is 14%, so one miss in the graph is not alarming.

The above discussion is written from a statistical viewpoint. An engineering viewpoint may reveal more. For example, if the estimates for the individual bins (plotted as dots in this section) drop very suddenly, it may be that the mechanism has changed. If the time bins correspond to plant age, the frequent early events may correspond to a learning period. Such conjectured causes should be investigated, and confirmed or rejected based on more detailed study of the events. As is typical, a statistical analysis only puts up road signs, pointing to interesting subjects for engineering investigations.

The statistical analysis is illustrated with an example here. The examples given earlier in this chapter could be used, but Example 2.1 is more interesting when investigating lack of fit.

Example 2.1 stated that a particular plant had 34 unplanned reactor trips while at power in 1987-1995. Table 7.3 gives the dates of those initiating events. This data set is a portion of the database used by Poloski et al. (1999a). This particular plant had its initial criticality on 1/3/87 and its commercial start on 5/2/87.

Table 7.3 Dates of initiating events at one plant, 1987-1995. (from Example 2.1)

01/21/87	04/03/87	06/17/87	11/08/87	02/07/89	07/15/92
01/22/87	04/12/87	06/21/87	03/09/88	02/22/89	07/17/92
02/27/87	04/14/87	06/22/87	10/14/88	03/14/89	10/12/95
03/11/87	04/21/87	07/09/87	10/30/88	10/09/89	11/05/95
03/13/87	04/22/87	08/04/87	01/16/89	06/03/91	
03/31/87	05/24/87	11/07/87	02/06/89	07/12/92	

These events were grouped by calendar year. Because reactor trips occur only when the reactor is at power, the relevant normalizing time is critical time, given in Table 7.4 as critical years.

The now-familiar picture is given in Figure 7.12. This figure shows that the first observed value, for 1987, is well above the fitted line, and the second observed value, for 1988, is well below the fitted line. In fact, the assumed model seems to try to force data with an L-shaped trend into a smooth exponentially decreasing trend.

It appears that the plant had a learning period, during which initiating events were quite frequent, followed by a period with a much smaller frequency. Examination of Table 7.4 shows that the learning period seems to have lasted until the summer of 1987 (three events in June, one each in July and August, and only infrequent events after that). It is not certain that the explanation is “learning” in the usual sense, but it is clear that the event frequency dropped suddenly about six months after the initial criticality.

Table 7.4 Initiating events and reactor-critical-years.

Calendar year	Initiating events	Reactor-critical-years
1987	19	0.70936
1988	3	0.75172
1989	6	0.79482
1990	0	0.89596
1991	1	0.81529
1992	3	0.75123
1993	0	0.99696
1994	0	0.82735
1995	2	0.83760

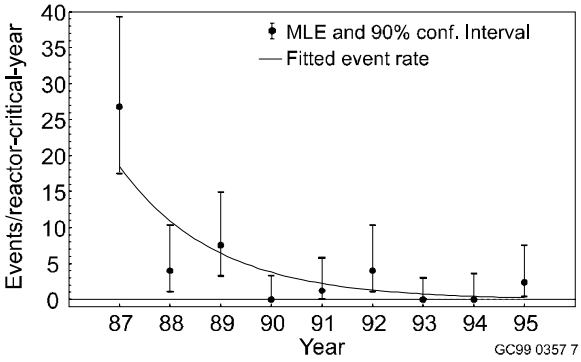


Figure 7.12 Annual frequency of initiating events, with fitted exponentially decreasing $\delta(t)$ and simultaneous 90% confidence band on $\delta(t)$.

Incidentally, a plot based on Bayesian calculations would show the same general information as Figure 7.12. Replace the MLE fitted line by the posterior median, and replace each confidence interval by a credible interval for δ based on a noninformative prior and one year's data. For an example of such a Bayesian plot, see Figure 7.18 in Section 7.4.3.

In ordinary least-squares fitting, it is standard to plot **residuals**, where each residual is defined as the observed value minus the fitted value. Under the assumed model, the residuals do not all have the same variance, so sometimes the **standardized residuals** are plotted, where a standardized residual is the residual divided by its theoretical standard deviation.

In the present context, the ith count, X_i, is assumed to be Poisson with mean $s_i\delta(t_i)$. The ith residual, denoted r_i, is

$$r_i = x_i - s_i\hat{\lambda}(t_i) .$$

The variance of a Poisson random variable equals the mean. Therefore, the standardized residual is

$$r_i / \sqrt{s_i\hat{\lambda}(t_i)} .$$

In the context of binned data, these are also sometimes called the **Pearson residuals** or the **chi-squared residuals**, because the sum of the squared Pearson residuals is equal to the Pearson chi-squared statistic. A plot of the these residuals against time may be helpful.

Figure 7.13 plots the standardized residuals against calendar year, for the example of Tables 7.3 and 7.4. This plot shows severe lack of fit. The standardized residuals should be approximately normal (0,1), and so should be mostly between !2 and 2. A value greater than 3.5 is just too large. The plot also shows something that may not be evident from Figure 7.12. The largest value corresponds to 1995, not 1987. This reflects the fact that in Figure 7.12, the 1995 confidence interval is farthest from the fitted line, in relative terms.

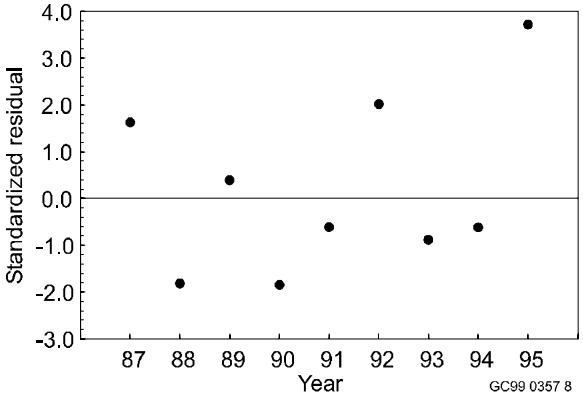


Figure 7.13 Standardized residuals, also called the Pearson chi-squared residuals, for data of Figure 7.12.

An informative plot for this example is the simple cumulative plot, introduced in Chapter 6 (see Figure 6.23). This plot would normally be used to check on whether the event occurrence rate is constant. The slope is the event rate, and a nonconstant slope corresponds to a departure from a straight line.

The cumulative event count is plotted against event date in Figure 7.14. In this example, the plot shows a clear nonconstant slope, and moreover, the form of the nonconstancy is shown: a very large rate (slope) during the first year, followed by a somewhat smaller rate, and then a very small rate in the last years.

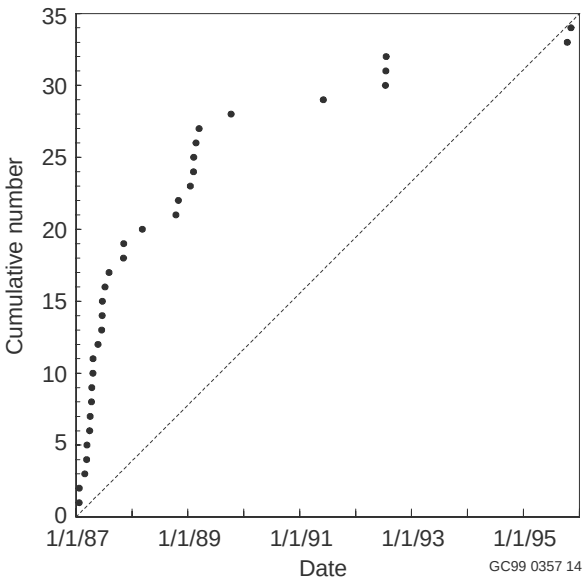


Figure 7.14 Cumulative count of initiating events, for data of Table 7.3 (Example 2.1).

Two comments must be made. First, this figure requires the unbinned data, which may not be available in every problem. Second, the use of calendar time instead of critical time on the horizontal axis may distort the figure somewhat.

If there is lack of fit to the exponential trend model, the analyst should try to identify the causes of the lack of fit. One frequent cause of lack of fit is systematic variation — the assumed form of the model is incorrect. In the examples of this section, systematic variation means that $\ln \delta(t)$ is not of the form $a + bt$. This was revealed by Figures 7.12 and 7.14. Another possible cause is extra-Poisson variance, additional sources of variation that are not accounted for in the Poisson model. Figure 7.13 may show this. The residual for 1995 is surprisingly large. Table 7.3 and Figure 7.14 both show that two events occurred in relatively quick succession in 1995, and that three events occurred in quick succession in 1992. If any of these events were dependent on each other, such dependence would exaggerate the normal random variation in the counts from year to year.

One must be careful about how to correct lack of fit. In this example, it is reasonable to delete the early history of the plant, the part corresponding to the learning period. This would flatten the fitted line for the rest of Figure 7.12, making it lower in 1988 and higher in 1995. Thus, dropping early data would make the late data fit better. One would have to perform the analysis to know whether this proposed solution will completely remove the lack of fit in the late years or only reduce it.

The above approach does not consist of throwing away data. Instead, it divides the data into two relatively homogeneous sets, which never should have been pooled. The set after the end of the learning period can be analyzed as described above. The other set, during the learning period, can also be analyzed. If it is thrown away, that is only because no one chooses to analyze it.

If dividing the data into homogeneous pieces does not correct the lack of fit, another option is to construct a more complex model. For example, δ could be modeled as a function of more variables than just t. Or one could postulate a random count with larger variance than the Poisson variance. Such topics are beyond the scope of this handbook.

7.2.6.2 Statistical Test for Goodness of Fit

7.2.6.2.1 Test Based on Poisson Maximum Likelihood

When Poisson maximum likelihood is used to fit the trend to the data, some software packages give two measures of goodness of fit, the Pearson chi-squared statistic and the **deviance**. The Pearson chi-squared statistic, denoted X^2, is the sum of squares of the Pearson residuals. The deviance is based on the theory of generalized linear models. It is defined as

$$D = 2\sum x_i \{\ln(x_i) \; ! \; \ln[s_i \, \hat{\lambda}(t_i)]\} \; .$$

For more details, see Atwood (1995) or books on the generalized linear model.

The assumed model is

$$H_0: \delta(t) = a + bt$$

for some (unknown) constants a and b. If this model is correct, and if the number of observations per bin is large, both X^2 and D have approximately a chi-squared distribution, with degrees of freedom equal to the number of bins minus the number of unknown parameters. In fact, the two statistics are asymptotically equal. For small samples, on the other hand, the two are not necessarily nearly equal to each other, nor is their distribution approximately chi-squared. The distribution of X^2 typically approaches the chi-square distribution faster than the distribution of D does.

These two statistics can be used to test whether H_0 is true. Four situations can arise in practice.

- Both X^2 and D are in the upper tail of the chi-squared distribution, larger than, say, the 95th percentile. This is evidence of lack of fit to the model, H_0. Report the p-value based on X^2. For example, if X^2 is at the 98th percentile of the chi-squared distribution, report a p-value of 0.02. Investigate the data to try to discover the reason for the lack of fit.
- Both X^2 and D are in the middle of the chi-squared distribution, say between the 5th and 95th percentiles. Then the model appears to fit adequately.
- Both X^2 and D are in the lower tail of the chi-squared distribution. This is an indication of overfit, with too complex a model to be justified by the data. Although such a situation will probably not arise with the two-parameter models of this chapter, it can arise when the model contains many factors, such as component age, manufacturer, environment, etc.
- X^2 and D are so different from each other that they give conflicting conclusions. That is, one statistic is in the upper tail of the chi-squared distribution and the other is in the lower tail, or one is in a tail and the other is in the middle. This can indicate one of two possibilities. (1) The data set may be too small to allow an accurate assessment of the goodness of fit. The problem often can be remedied by pooling the data to some extent. For example, it is possible to fit a loglinear model using nine one-year bins from Table 7.4. If X^2 and D conflict, try pooling the data into two-year bins, and so forth. (2) H_0 may be false in a way that one statistic detects and the other does not. X^2 and D are asymptotically equal only if H_0 is true, not if H_0 is false. In this case, if it is really important to decide whether H_0 should be rejected, one could try simulating data from the fitted model, and seeing what fraction of the simulated data sets produce a simulated X^2 or D as large as actually observed. That fraction would approximate the exact p-value, without relying on the asymptotic chi-squared approximation.

For the HPCI unplanned demand data in Table 7.1, the loglinear model seems to fit well. The values of X^2 and D are 4.90 and 5.12, respectively. These are both in the middle of a chi-squared distribution. The degrees of freedom, 5, equals the number of bins, 7, minus the number of unknown parameters, a and b.

For the initiating-event data of Table 7.4, X^2 and D are 28.93 and 24.17. These are both far out in the right tail of a chi-squared distribution with 7 degrees of freedom. This is very strong evidence against the loglinear model.

The results of these statistical tests are consistent with the conclusions based on the graphs.

7.2.6.2.2 Test Based on Weighted Least-Squares Fit

Consider fitting a function of the form

$$y = a + bt$$

based on observations y_i at times t_i. In the present context, y equals $\ln(x_i / s_i)$. In Section 7.2.4.6, the weighted sum of squares

$$\sum w_i[y_i - (a + bt_i)]^2$$

was minimized, with the weights equal to the inverses of the estimated variances of $Y_i = \ln(X_i/s_i)$. If the model assumptions are correct, and if the Y_is are approximately normally distributed, the weighted sum of squares has approximately a chi-squared distribution. The degrees of freedom d is the number of bins minus the number of unknown parameters. The degrees of freedom are 5 in the HPCI-demand example and 7 in the initiating-events example. (Purists will note that the chi-squared distribution applies if the weights are fixed in advance, not derived from the random data. This slight departure from theory is commonly ignored.)

If the weighted sum of squares is in the right tail of the chi-squared distribution, such as beyond the 95th percentile, this is evidence of lack of fit. As mentioned above, one common cause of lack of fit is systematic variation — the assumed form of the model is incorrect. In the examples of this section, that means that $\ln \delta(t)$ is not of the form $a + bt$. Another possible cause is extra-Poisson variance, additional sources of variation that are not accounted for in the Poisson model. To gain insight as to which contributors to lack of fit seem to be present, examine plots similar to Figures 7.12 through 7.14.

7.3 Unbinned Poisson Data

Example 6.6 and Table 7.3 are typical examples of the type of data considered here. That is, the exact event times are used. The corresponding summary tables of counts, given in Tables 7.1 and 7.4, are not used.

In principle, the exact event times contain more information than the summaries of counts in bins. The counts can be calculated from the exact event times, but

the exact event times cannot be retrieved from the count totals. Therefore, in principle, better procedures can be squeezed out of the exact event times. However, software based on binned event counts is more widely available. Also, little information is lost by grouping the event times into bins, unless the bins are very large and coarse. Therefore, use of binned data is usually the most practical method for the data analyst.

7.3.1 Bayesian Analysis

When the exact event times are used, it is difficult to write out the likelihood. In principle, one must write the likelihood of each event time, conditional on the previous times, and finally the probability of no events after the final observed event until the end of the observation period. [See Section 3.3 of Cox and Lewis (1966).] Formulas are given for several cases by Atwood (1992), but they are not intuitive.

Binning the data, as in Section 7.2, is much simpler. Moreover, the bins may be made as numerous and small as the analyst desires. Many of the bins would then have event counts of zero, but that is allowed. This approach would capture virtually all of the information in the data. In practice there is little advantage in constructing very fine bins, but the analyst who was intent of squeezing every last bit of information from the data could do it.

7.3.2 Frequentist Analysis

Frequentist analysis is also simpler when the data are binned, although Atwood (1992) works out the formulas for the MLEs and approximate confidence intervals for several cases that use the exact event times. The unified notation in that article does not make the expression immediately obvious for any particular model. Typically, the MLEs must be found through numerical iteration rather than directly from algebraic formulas.

The simplest approach is to bin the data and use the methods of Section 7.2. Remember that the bins must not be too fine; a conservative rule of thumb says that most of the bins should have expected event counts of five or more.

The exception — the only relatively easy case with unbinned data — is the power-law model when the process is observed from time zero. Typically, the data collection begins at some time in the middle of operation, but in those rare cases when the data collection starts at time zero and ends at some time $\mathcal{J}$, the MLEs of the parameters in the two parameterizations of the power-law model are:

$$\hat{b} = n / \sum_{i=1}^{n} \ln(\tau / t_i),$$

$$\hat{c} = \tau / n^{1/\hat{b}}, \text{ and}$$

$$\hat{A} = n\hat{b} / \tau^{\hat{b}} .$$

Here n is the number of events, t_i is the time of the ith event, and $\mathcal{J}$ is the final time in the data-observation period. These formulas can be obtained by translation into the notation of Section 7.2.2.1 of formulas in Bain and Engelhardt (1991, Chap. 9, Eq. 13) or Atwood (1992, Section 6.1). Those references also consider confidence intervals and goodness-of-fit tests.

7.4 Binomial Data

This section parallels Section 7.2 closely. Only the formulas are different, because the section deals with failures on demand instead of events in time. Because of the similarity to Section 7.2, some of the topics are given a cursory treatment here. These topics are completely analogous to the material in Section 7.2, where a fuller description can be found.

7.4.1 Examples

Example 6.10 can be used. This example consisted of 63 demands for the HPCI system during 7years, and 12 failures. It is convenient to combine the data into bins, such as calendar months, calendar years, etc. Such binning summarizes the data in a compact way. For goodness-of-fit tests, discussed in Section 7.4.6, binning is not merely convenient – it is required. If the bins are too fine (too few failures and successes expected in each bin) then the goodness-of-fit statistics X^2 and D will be inconsistent with each other, and neither will have an approximate chi-squared distribution under H_0. On the other hand, the bins must not be too coarse. As in Section 7.2, denote the midpoint of the ith bin by t_i. The bins must be small enough so that the expected number of failures in the bin can be approximated by the number of demands in the bin times $p(t_i)$.

The data from Example 6.10 are summarized by calendar year in Table 6.14, which is repeated here as Table 7.5.

In this example the bins correspond to calendar years. Other examples could be constructed in which the bins

correspond to ages, so that p would be modeled as a function of age rather than calendar time.

Table 7.5 HPCI failures on demand, by year (from Example 6.10).

Calendar year	Failures	Demands
1987	4	16
1988	2	10
1989	1	7
1990	3	13
1991	2	9
1992	0	6
1993	0	2

7.4.2 Model

7.4.2.1 General Model

The model is the same as that in Sections 2.3.1 and 6.3, except now the probability p depends on time, t. Thus, the model assumptions are:

1. The outcome of a demand at time t is a failure with some probability $p(t)$, and a success with probability $1 - p(t)$.

2. Occurrences of failures on different demands are statistically independent.

The number of demands and their times are assumed to be fixed, and the outcome on each demand is assumed to be random.

7.4.2.2 Logit Model

By far the most commonly used functional form for $p(t)$ is the **logit** model. In this model the logit transform of $p(t)$ is a function of unknown parameters. Fitting such a model to data is sometimes called **logistic regression**. The model that will be used in this section is

$$\text{logit}[p(t)] = a + bt ,$$

where the **logit** function is defined as

$$\text{logit}(p) = \ln[\, p/(1 - p)\,] .$$

This function was also encountered in Section 6.3.2.5.2 and in Appendix A.7.9, where the logistic-normal distribution is introduced. Like the loglinear model for $\mathcal{E}(t)$, the logit model for $p(t)$ is a model in the class of generalized linear models.

A frequently used relation is that

$$y = \text{logit}[p(t)] \;/\; \ln\{\, p(t)/[1 - p(t)]\,\}$$

is equivalent to

$$p(t) = \text{logit}^{-1}(y) \;/\; e^{y}/(1 + e^{y}) , \qquad (7.7)$$

denoting the inverse function of the logit by logit^{-1}. Figure 7.15 shows $\text{logit}[p(t)]$ as a function of t, and Figure 7.16 shows $p(t)$ itself as a function of t. Notice that in Figure 7.16 the value of $p(t)$ stays between 0.0 and 1.0, as it should.

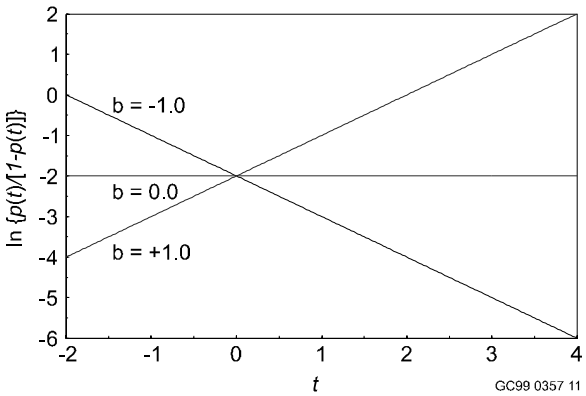


Figure 7.15 Plot of $\ln\{\, p(t)/[1 - p(t)]\,\} = a + bt$, with $a = -2$ and three values of b.

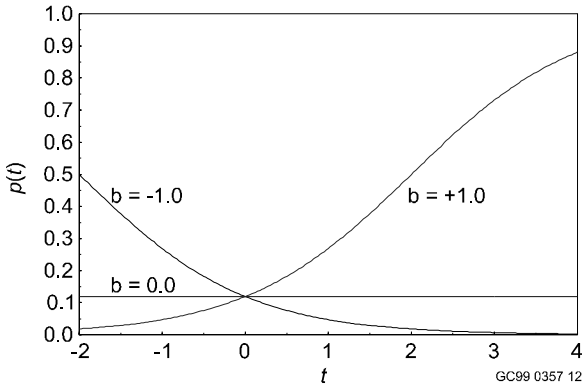


Figure 7.16 Plot of $p(t)$ as a function of t, corresponding to Figure 7.15.

The parameters have simple interpretations: a is the value of the logit of p when $t = 0$, and b is the slope of the logit of p.

7.4.2.3 Loglinear Model

If p is small, then logit(p) is close to ln(p), and the logit model could be approximated by

$$\ln[p(t)] = a + bt \ .$$

This is a loglinear model, just as in Section 7.2.2.2. Software programs for analyzing a generalized linear model always include the logit model as one special case, and software programs for logistic regression are based on the logit model. However, if for some reason the analyst has software that only allows for the log-transformation, not the logit-transformation, that software is probably adequate as long as p is small.

7.4.3 Bayesian Estimation with Logit Model

The large-sample theory mentioned in Section 7.2.3 applies here as well. As the data set becomes large (many demands and failures) the form of the likelihood function approaches the form of a normal density for the two variables a and b. Therefore, with large data sets the **conjugate** prior is normal: if a and b are given a normal prior distribution, the posterior will be approximately normal, with very good approximation as the data set becomes large. The corresponding **noninformative** prior for a and b is the limiting case as the variance approaches infinity, which is a constant density.

For the work here, it will be assumed that a computer program produces a sample from the posterior distribution. The theory sketched above will then lead us to the conclusions that the posterior distributions not only *appear* normal, they really *are* normal, or very close to normal.

A convenient, and free, software package is BUGS (1995), Bayesian inference Using Gibbs Sampling. This was also used in Section 7.2.3, where it is described in more detail. Example 6.10 is analyzed here using the Windows version, WinBUGS, assuming a logit model.

Using the data of Example 6.10 (Table 7.5), BUGS was used to model logit$p(i) = a + bi$, for i from 1 to 7. Then $X(i)$ was modeled as having a binomial($n(i)$, $p(i)$) distribution, where $n(i)$ is the number of demands in year i. Finally, a and b were given very diffuse prior normal distributions.

Figure 7.17 shows the BUGS script that was used to analyze the data.

```
model
{
   for (i in 1:m) {
      p[i] <- exp(a + i*b)/(1 + exp(a+i*b))
      x[i] ~ dbin(p[i], n[i])
   }
   a ~ dnorm(0.0, 0.0001)
   b ~ dnorm(0.0, 0.0001)
}
```

Figure 7.17 BUGS script for analyzing data of Example 6.10.

This uses the Equation 7.7 for expressing logit(p) in terms of the normally distributed quantity $a + ib$. Note the way BUGS happens to parameterize distributions, putting p before n in the list of binomial parameters, and parameterizing the normal distribution in terms of the precision $J = 1/F^2$. The reason for using precision is explained in Section 6.6.1.2.1. A precision of 0.0001 in the script corresponds to a standard deviation of 100. That gives a very diffuse distribution.

The script was executed with four separate starting values of a and b, generating four **chains** of values, each 10,000 elements long. The first 500 elements of each chain were discarded, and the remaining 38,000 elements were used to estimate the posterior distributions. Table 7.6 summarizes the posterior distributions for a and b. When interpreting these summaries, be aware that a and b are not independently distributed.

Table 7.6 Posterior statistics for a and b, for loglinear model of Example 6.5.

	a	b
mean	!0.8838	!0.2085
median	!0.8654	!0.204
st. dev.	0.6477	0.1961
5th percentile	!1.981	!0.5395
95th percentile	0.1471	0.1063

The table shows that the mean and median are nearly equal in each case, and the 5th and 95th percentiles are approximately 1.645 standard deviations from the mean. This approximation is poorest for a, with !0.8838 + 1.645 × 0.6477 = 0.182, some-

what larger than the reported 95th percentile. That is, the distributions appear to be approximately normal, as anticipated by the theory mentioned earlier, but the approximation is not perfect in the tails.

The mean of *b* is negative, but not strongly so, only 1.06 standard deviations below 0. Treating *b* as normally distributed, a table of the normal distribution shows that *b* is negative with probability 0.86 and positive with probability 0.14. Therefore, we are not really sure that the trend is downward.

The posterior distribution of *p* is approximately logistic-normal in any year. It is shown in Figure 7.18. The median is plotted as a solid line and the 5th and 95th percentiles are shown as dashed lines. This is analogous to Figure 7.7 for *8*.

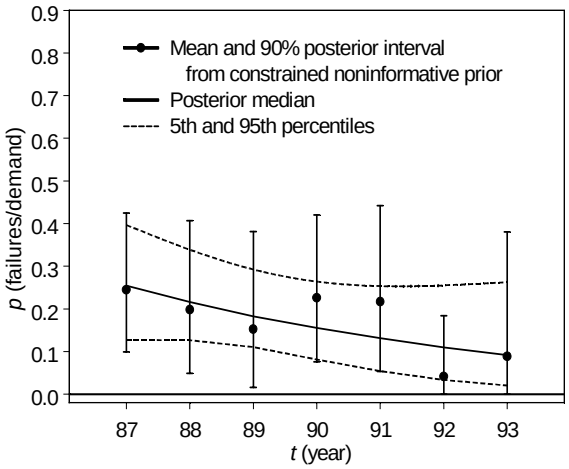


Figure 7.18 Posterior trend line for *p* with 90% credible band, for data of Table 7.5. In addition, annual estimates and 90% credible intervals are shown, based on constrained noninformative prior.

Figure 7.7 also plotted the MLEs, based on each year's data, as dots. To illustrate the graphical possibilities, Figure 7.18 is constructed somewhat differently. The total data set has 12 failures in 63 demands. Therefore we constructed the constrained noninformative prior with mean 12/63. Interpolation of Table C.8 shows that this prior is approximately beta(0.324, 1.376). For each year of data, this prior was updated to obtain the posterior for that year; the 90% credible interval was plotted as a vertical line, with a dot showing the posterior mean.

All the intervals overlap the fitted trend line. This is graphical evidence that the logit model fits the data well. (Many other models might also fit this sparse data set well.)

7.4.4 Frequentist Estimation with Logit Model

7.4.4.1 Point Estimation

The model is analyzed by finding the MLEs of *a* and *b*, based on binomial counts. The discussion below will sometimes call this the **binomial-maximum-likelihood** method. The software finds the estimates by numerical iteration.

When this model is fitted to the data of Table 7.5, a fitted trend is found, which can be overlaid on Figure 6.38. It is not shown here, but will be displayed with a simultaneous confidence band in Section 7.4.4.4.

7.4.4.2 Confidence Intervals for *a* and *b*

As in Section 7.2.4.2, almost all software for estimating *a* and *b* reports standard errors, estimates of the standard deviations of the maximum likelihood estimators. The estimators are assumed to be approximately normally distributed, which is valid unless the sample size is small.

Therefore, as in Section 7.2.4.2, a 100(1 ! ")% confidence interval for *b* is

$$\hat{b} \pm z_{1-\alpha/2} se(b)$$

where $\hat{b}$ is the estimate, and $z_{1 ! "/2}$ is the 1 ! "/2 quantile of the normal distribution. The term *se*(*b*) is the standard error of *b*, the estimated standard deviation of the estimator.

The confidence interval for *a* is similar.

7.4.4.3 Test for Presence of Trend

Consider the two hypotheses defined by:

H_0: $p(t)$ is constant.
H_1: $p(t) = \text{logit}^{!1}(a + bt)$, $b \ldots 0$.

Note, the null hypothesis H_0 is true if $p(t) = \text{logit}^{!1}(a + bt)$ and *b* is zero. Therefore, with this choice of an alternative hypothesis, the test of H_0 is the same as a test that $b = 0$, the test given above based on a confidence interval for *b*.

As in Section 7.2.4.2, the hypothesis

H_0: $b = 0$,

the hypothesis of no trend, is rejected at level 0.10 if the 90% confidence interval for b is entirely on one side of 0. The hypothesis is rejected at level 0.05 if the 95% confidence interval is entirely on one side of 0, and so forth. Most software packages print out a significance level at which the hypothesis that $b = 0$ is rejected, the p-value for the trend.

This is different from the tests of Chapter 6. Section 6.3.3.2.2 uses a different alternative hypothesis,

H_1: $p(t)$ is not constant .

It also uses a different test, the chi-squared test. Section 6.3.3.2.2 commented that the test is not very powerful against the alternative of a trend in p, because it considers such a broad class of possibilities as the alternative hypothesis.

Consider Example 6.10, with the HPCI failures during unplanned demands, as summarized in Table 7.5. In Section 6.3.3.2.2, the chi-squared test rejected the hypothesis of constant p with p-value 0.77. That is, the test found no evidence of nonconstant p. The present test of $b = 0$ rejects this hypothesis with p-value 0.30. That is, this test still does not reject the hypothesis of constant p. However, it notices the slightly falling values of p in Figure 6.38, and therefore sees somewhat stronger evidence against constant p than the chi-squared test did. Incidentally, the test based on $\hat{b}$ and the Wilcoxon-Mann-Whitney test for trend (Section 6.3.3.2.2) reach very similar conclusions.

7.4.4.4 Confidence Intervals and Confidence Bands

Confidence intervals for $p(t)$ at a particular t and simultaneous confidence bands valid for all t both are based on the approximate normality of the MLEs $\hat{a}$ and $\hat{b}$. The software finds an approximate $100(1 - \alpha)$% confidence interval for logit$[p(t)]$ as

$$\hat{a} + \hat{b}t \pm [z_{1-\alpha/2} \times se(a + bt)] \qquad (7.8)$$

where, as before, $z_{1-\alpha/2}$ is the $100(1 - \alpha/2)$ percentile of the standard normal distribution, and $se(a + bt)$ is the standard error, the estimated standard deviation of $\hat{a} + \hat{b}t$. The standard error depends on the value of t, and accounts for the fact that the MLEs $\hat{a}$ and $\hat{b}$ are statistically correlated, not independent. The confidence interval for $p(t)$ itself is found by inverting the logit function. If L and U are the lower and upper confidence bounds for logit$[p(t)]$, that is, for $a + bt$, then

$$\text{logit}^{-1}(L) = e^L/(1 + e^L) \text{ and}$$
$$\text{logit}^{-1}(U) = e^U/(1 + e^U) \qquad (7.9)$$

are the corresponding confidence bounds for $p(t)$. Manipulation of Equation 7.8 allows the analyst to convert from one degree of confidence to another, say from 90% to 99%, by using different percentiles of the normal distribution and the single standard error found by the software.

As discussed in Section 7.2.4.5, a confidence interval is valid at one t, and the band constructed from the individual confidence intervals is not simultaneously valid for all t. A simultaneous $100(1 - \alpha)$% confidence band for logit$[p(t)]$ is found by replacing $z_{1-\alpha/2}$ in Equation 7.8 by

$$[\chi^2_{1-\alpha}(r)]^{1/2} ,$$

with r equal to the number of estimated parameters, 2 in Equation 7.8. This is exactly the value that was used in Section 7.2.4.5. The only difference is that there the confidence band was for $\ln\delta(t)$ and here it is for logit$[p(t)]$. The confidence band for $\delta(t)$ was found by inverting the logarithm function, that is, by taking an exponential. The confidence band for $p(t)$ is found by inverting the logit function: if L and U are now used to denote the lower and upper edges of the simultaneous confidence band for logit$[p(t)]$ at some t, the corresponding points on the confidence band for $p(t)$ are given by Equation 7.9.

Figure 7.19 shows the data from Table 7.5, plotted as in Figure 6.38 but now with the fitted trend line and the simultaneous 90% confidence band overlaid. The confidence band can easily contain a horizontal line; this is consistent with the fact that the hypothesis of constant p cannot be rejected.

7.4.4.5 Alternative Using Least-Squares Software

The model assumes that logit$[p(t)] = a + bt$. Therefore, as in Section 7.2.4.6, one might decide simply to use least-squares software as follows. First, estimate p for each bin, based on only the data for that bin:

$$\hat{p}_i = x_i / n_i .$$

Then fit logit($\hat{p}_i$) to $a + bt_i$ by least squares. The same problems that were mentioned in Section 7.2.4.6 must be addressed here.

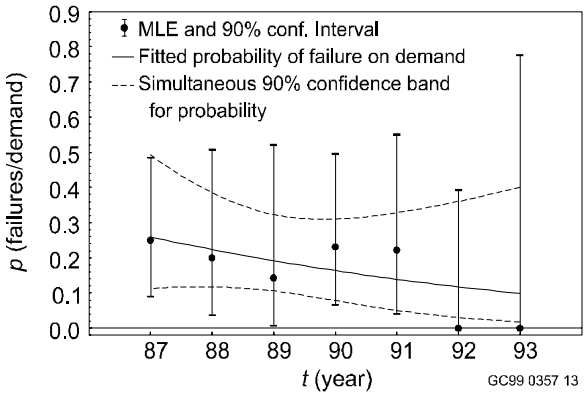


Figure 7.19 Annual estimates of p, fitted trend line and 90% confidence band for $p(t)$, for data of Table 7.5.

First, if any observed failure count, x_i, equals either 0 or the demand count n_i, the MLE $\hat{p}$ will be 0.0 or 1.0 for that bin, and the logit will be undefined. In the data of Table 7.5, this happens for the final two years. The following ways around this have been proposed, analogues of proposals in Section 7.2.4.6:

1. Instead of estimating p_i by x_i/n_i, use $(x_i + ½)/(n_i + 1)$. This is equivalent to replacing the MLE by the posterior mean based on the Jeffreys noninformative prior.
2. Estimate p_i by the posterior mean based on the constrained noninformative prior. In this case, the constraint could be that the prior mean equals the observed overall mean $\Sigma_i x_i/\Sigma_i n_i$, or equals a modification to guarantee a positive number, $(\Sigma_i x_i + ½)/(\Sigma_i n_i + 1)$.

Such ways tend to reduce the trend slightly, because they add a constant to all the failure and demand counts, slightly flattening out any original differences.

The second point that must be considered is that the variance of the estimator of p is not constant. Therefore, the following iteratively reweighted least-squares method can be used. Assume that p in the ith bin is estimated by $(X_i + \alpha)/(n_i + \alpha + \beta)$. If the simple MLE is used, then α and β are both zero. If method 1 above is used, then $\alpha = ½$ and $\beta = ½$. If method 2 above is used, then α and β must be found from Table C.8. Neter and Wasserman (1974, Eq. 9.51) state that the asymptotic variance of logit(MLE of p_i) is

$$1/[n_i\, p_i(1 - p_i)] .$$

The method given here is a generalization when α and β are not both zero, setting the weight w_i to the inverse of the asymptotic variance of the estimator.

Begin by assuming that p is constant, and let $\hat{p}_i$ be some simple estimate of p, the same for all i. Fit logit$[(x_i + \alpha)/(n_i + \beta)]$ to a straight line with weighted least squares, and weights

$$w_i = \frac{(n_i \hat{p}_i + \alpha)^2 [n_i(1 - \hat{p}_i) + \beta]^2}{n_i \hat{p}_i (1 - \hat{p}_i)(n_i + \alpha + \beta)^2} .$$

Calculate the resulting estimates of p_i,

$$\hat{p}_i = \text{logit}^{-1}(\hat{a} + \hat{b}t_i) .$$

Recalculate the weights with these estimates, and refit the data to a straight line using weighted least squares. Repeat this process until the estimates stabilize.

The third and final point is that least-squares fitting typically assumes that the data are approximately normally distributed around the straight line. In the present context, this means that logit($\hat{p}_i$) is assumed to be approximately normally distributed. This assumption is acceptable if the number of failures in each bin is not close to zero or to the number of demands. The variance of the normal distribution is then estimated from the scatter around the fitted line. This differs from typical treatment of binomial data, where the mean determines the variance.

A 90% confidence interval for $p(t)$ at a particular t is given by

$$\text{logit}^{-1}\{\hat{a} + \hat{b}t \pm [t_{0.95}(d) \times se(a + bt)]\} \qquad (7.10)$$

where $t_{0.95}(d)$ is the 95th percentile of Student's t distribution with d degrees of freedom, just as with $\delta(t)$ in Section 7.2.4.6. The software will report the value of d. It is the number of bins minus the number of estimated parameters, 7 − 2 in the example of Table 7.5. The form of this equation is very similar to other equations in Sections 7.2 and 7.4, although the estimates and standard deviation are calculated somewhat differently.

A simultaneous 90% confidence band has the same form, but the multiplier $t_{0.95}(d)$ is replaced by

$$[2F_{0.90}(r, d)]^{½} ,$$

where $F_{0.90}(r, d)$ is the 90th percentile of the F distribution with r and d degrees of freedom.

7.4.5 Comparison of Bayesian and Frequentist Estimation with Logit Model

When the Bayesian analysis uses very diffuse priors, the conclusions of the two analyses will be numerically similar.

The posterior median in Figure 7.18 is very close to the fitted line (the MLE) in Figure 7.19. The 90% credible band in Figure 7.18 is narrower than the simultaneous confidence band in Figure 7.19, because the simultaneous confidence band is based on an inequality. It would be close to the frequentist bounds that are valid at any one t, if such a graph were calculated. The vertical lines in Figure 7.18, representing credible intervals for p based on individual years of data, are generally close to the confidence intervals in Figure 7.19, except for the years with little data, 1992 and 1993. For those two years, the confidence intervals are quite wide, but the credible intervals are shorter, under the influence of the prior mean of 0.19.

Frequentist estimation relies on approximate normality of the estimators, and therefore does not work well with data having few observed events. Bayesian estimation obtains normal posteriors when the data set is large, but does not fail entirely when the data set is small – it merely obtains a different, non-normal, posterior.

Most frequentist software packages for analyzing trends include calculations for investigating the goodness of fit of the model, as will be seen in Section 7.4.6. Current Bayesian software may neglect this issue of model validation.

7.4.6 Model Validation

The two assumptions for a time-dependent binomial process are given at the beginning of Section 7.4.2. The first assumption is difficult to test from data. The other assumption is that outcomes on distinct demands are independent. One kind of dependence is serial dependence. Positive serial dependence means that failures tend to be followed by more failures, for example if a failure is misdiagnosed the first time, or if a single cause results in a number of failures before it is corrected. Negative serial dependence means that failures tend to be followed by successes, for example if the major cause of failure is wearout, at which time a new component is installed (without any failures from installation problems).

Positive serial dependence results in failures tending to cluster together, with relatively long gaps between failures. When the data are collected into bins, this can translate into unusually high or low event counts in individual bins. This can be discovered by goodness-of-fit tests, considered below. However, it is impossible to decide, from the failure counts alone, whether the outcomes are serially correlated or whether p is going up and down. The cause can be determined only by an investigation to discover the failure mechanisms.

A negative serial dependence results in less-than-expected variation in the event counts. A goodness-of-fit test will report a p-value near 1.0, indicating surprisingly good fit, too good to be believable.

The final assumption made when fitting a trend model is the form of the model. Goodness-of-fit tests are designed for testing this assumption. In fact, a goodness-of-fit test is an all-purpose test for the various assumptions, although it is not good at deciding which assumption may be violated.

7.4.6.1 Graphical Check for Goodness of Fit

The natural graphical check is to compare the observed values to the fitted line.

Figure 7.18 and 7.19 each show such a plot for the data of Table 7.5 (Example 6.10). Either figure may be used. In Figure 7.18, each credible interval overlaps the fitted trend line, and in Figure 7.19, the 90% confidence interval for each year overlaps the fitted trend line. Because no year deviates strongly from the overall trend, the data appear consistent with the assumption of the logit model.

The discussion at the end of Section 7.2.6.1 applies here as well, concerning interpretation of a few intervals' failure to overlap the fitted line, and concerning the need for an engineering assessment of any strange patterns.

As with the loglinear model for $\delta(t)$, the residuals and standardized residuals can be plotted for $p(t)$. Software may report these as the "raw residuals" and the "Pearson chi-squared residuals," respectively.

In the present context, the ith count, X_i, is assumed to be binomial with mean $n_i p(t_i)$. The ith residual, denoted r_i, is

$$r_i = x_i - n_i \hat{p}(t_i) .$$

The variance of a binomial(n, p) random variable equals $np(1 - p)$. Therefore, the standardized residual is

$$r_i / \sqrt{n_i \hat{p}(t_i)[1 - \hat{p}(t_i)]} \ .$$

These are also sometimes called the Pearson residuals or chi-squared residuals. A plot of the standardized residuals against time may be helpful, as in Section 7.2.6.1.

Also, a simple cumulative plot may be informative, as it was in Section 7.2.6.1. In the present example, the cumulative plot is given in Figure 6.39, which shows no pattern of interest. In other data sets, such a plot might not only show nonconstancy in p, but it might suggest the form of the nonconstancy.

The cumulative plot shows failures per demand, revealing changes in p as a function of the demand count. However, if the rate of demands is changing, as it is in the present example, the plot can give a distorted picture if p is regarded as a function of calendar time. When p is constant, the issue of distortion is irrelevant — a constant p is constant, whether it is regarded as a function of demand count or of calendar time.

Just as in Section 7.2, lack of fit may be caused by systematic variation or by extra-binomial variance, additional sources of variation that are not accounted for in the binomial model. See the discussion in Section 7.2.6.1.

7.4.6.2 Statistical Test for Goodness of Fit

7.4.6.2.1 Test Based on Binomial Maximum Likelihood

Just as in Section 7.2, software packages that use binomial maximum likelihood may give two measures of goodness of fit, the Pearson chi-squared statistic, denoted X^2, and the deviance, denoted D. The discussion in Section 7.2.6.2.1 applies here as well.

For the HPCI failure data in Table 7.5, the logit model seems to fit well. The values of X^2 and D are 2.12 and 2.91, respectively. These are both in the middle of a chi-squared distribution with five degrees of freedom. The degrees of freedom, five, equals the number of bins, seven, minus the number of unknown parameters, a and b. The p-value for lack of fit, based on X^2, is 0.84, indicating very good fit.

7.4.6.2.2 Test Based on Weighted Least-Squares Fit

Consider fitting a function of the form

$$y = a + bt$$

based on observations y_i at times t_i. In the present context, y equals logit(x_i/n_i). As in Section 7.2, if the model assumptions are correct, and if the Y_is are approximately normally distributed, the weighted sum of squares has approximately a chi-squared distribution. The degrees of freedom d is the number of bins minus the number of unknown parameters. The number of degrees of freedom is 5 in the HPCI-failure example. The discussion in Section 7.2.6.2.2 applies here as well.

7.5 Discussion

This section ties together some specific methods given in Chapters 6 and 7, showing the unifying formulations. Readers who are happy simply using the earlier recipes may skip this section.

7.5.1 Generalized Linear Models

Some software packages that implement the loglinear and logit models do so in the framework of the generalized linear model. Such models are described in a highly mathematical way by McCullagh and Nelder (1989), and in an introductory way in Appendix B-2 of Atwood (1995). This model has several elements, a **random component**, a **systematic component**, and a **link**, a function relating the random and the systematic components.

- The random component consists of some independent observations $Y_1, \ldots, Y_m$. This is thought of as an m-dimensional vector, $\mathbf{Y}$. The examples of this chapter have been the normalized Poisson event count, $Y_i = X_i/s_i$, and the fraction of binomial failures on demand $Y_i = X_i/n_i$.
- The systematic component is an m-dimensional vector $\boldsymbol{\theta}$, with the ith element θ_i related to explanatory variables, and to unknown parameters in a linear way. The example of this chapter has been $\theta_i = a + bt_i$, where a and b are unknown parameters and the explanatory variable t_i is the calendar time or age for Y_i.
- The link is a function g with

$$\theta_i = g[E(Y_i)] \ . \qquad (7.11)$$

In this chapter, the links have been the log function for Poisson data and the logit function for binomial data.

Thus the two examples of this chapter have been the Poisson example, with

$$\ln[E(X_i/s_i)] = \ln[\lambda(t_i)] = a + bt_i ,$$

and the binomial example, with

$$\text{logit}[E(X_i/n_i)] = \text{logit}[p(t_i)] = a + bt_i .$$

This is the terminology used by many statistical software packages. The analyst must specify the distribution for the random component, the form of the systematic component in terms of the unknown parameters and the explanatory variables, and the link function.

Other software packages use a slight variant of this terminology. This is a generalized linear model with an **offset**, replacing Equation 7.11 by

$$g[E(Y_i)] = \beta_i + \text{offset}_i . \qquad (7.12)$$

In the Poisson example, let Y_i be the Poisson count itself, X_i, not the normalized count X_i/s_i. Then the expected value of Y_i is $s_i\lambda(t_i)$, with $\ln\lambda(t_i)$ modeled as $a + bt_i$. To satisfy Equation 7.12, let the offset term be $\ln(s_i)$. Then we have:

$$\begin{aligned} g[E(Y_i)] &= \ln[E(X_i)] \\ &= \ln[s_i\lambda(t_i)] \\ &= \ln[\lambda(t_i)] + \ln(s_i) \\ &= a + bt_i + \ln(s_i) \\ &= \beta_i + \text{offset}_i . \end{aligned}$$

In this version of the model, the software package requires the analyst to specify the distribution of the random component, the form of the systematic component, the link function, and the offset. The disadvantage of this formulation is the extra term that must be specified, the offset. The advantage is that the distribution of X_i is Poisson, whereas the distribution of X_i/s_i is hard to specify because it does not have a standard name.

Much more elaborate models can be constructed in this general framework, by adding more explanatory variables. For example, both calendar time and age of the individual component could be treated together in one model as explanatory variables. The explanatory variables do not even have to be continuous. Manufacturer, system, and plant could be used as discrete explanatory variables. The possibilities are limited only by the availability of data. However, such models go beyond the limited scope of this handbook.

7.5.2 The Many Appearances of the Chi-Squared Test

In Chapter 6, the Pearson chi-squared test was used to test whether λ or p was constant. In Chapter 7, λ or p is assumed to be nonconstant, yet the chi-squared test is used anyway. Also, the chi-squared test was used in Section 6.6.2.3.2 to test whether durations had an assumed distributional form.

To see the unity in this apparent diversity, note first that in every case the Pearson chi-squared statistic has the form

$$X^2 = \Sigma_i \, (\text{observed}_i - \text{expected}_i)^2 / \text{expected}_i .$$

To clarify possible confusion one must think about the hypothesis being tested. The big general theorem states that when H_0 is true, X^2 has approximately a chi-squared distribution, and the degrees of freedom is the number of unknown parameters under H_1 minus the number of unknown parameters under H_0. The approximation is valid when the degrees of freedom stays constant and the size of the data set becomes large. This is now applied to the specific cases in this handbook.

Consider first Poisson data with event rate λ. In Chapter 6, the null and alternative hypotheses were:

H_0: λ is constant.
H_1: λ is not constant.

The data were partitioned into c cells. These cells may have corresponded to different sources of data, such as different plants, or they may have resulted from binning the data, for example, corresponding to c years. If H_0 is true there is one unknown parameter, the single value of λ. If, instead, H_1 is true, there are c unknown parameters, the values of λ in the different cells. Therefore, by the big general theorem, when H_0 is true X^2 is approximately chi-squared with $c - 1$ degrees of freedom. This is the result stated in Section 6.2.3.1.2.

Consider now the corresponding case with binomial data. The data fall into a $2\times J$ contingency table. The value J corresponds to the J sources of data or J bins, and the two rows correspond to failures and successes. The null and alternative hypotheses were:

H_0: p is constant
H_1: p is not constant .

If H_0 is true, there is one unknown parameter, p. If, instead, H_1 is true there are J unknown parameters, corresponding to the J bins or data sources. Therefore, the big general theorem says that the degrees of freedom for X^2 is $J - 1$, just as stated in Section 6.3.3.1.2.

Consider now the setting of Section 7.2, with $\lambda(t)$ modeled as $a + bt$. The time axis was divided into c bins. To test the goodness of fit, the hypotheses were:

H_0: $\lambda(t) = a + bt$,
H_1: λ for each bin is arbitrary.

Under H_0 there are two unknown parameters, a and b. Under H_1, the number of unknown parameters is the number of bins, since each can correspond to a different λ. Therefore, the big general theorem says that the degrees of freedom, when testing H_0, is $c - 2$. This agrees with Section 7.2.6.2. Recall that the chi-squared distribution is approximate, for large data sets. The deviance was used as a backup check, to help ensure that the data set was large enough.

The treatment of p in Section 7.4.6.2 is exactly parallel to that of λ.

Finally, the chi-square test was used in Section 6.6.2.3.2 to test whether durations follow a distribution of an assumed form. A duration was denoted as T. To be specific, consider a case in which the assumed distribution of T had two unknown parameters. Possibilities include the lognormal(ξ, σ) distribution and the gamma(α, β). The hypotheses in this setting were:

H_0: T has a distribution of the assumed form.
H_1: T has some other distribution.

The data were partitioned into c bins, so that every observed value of T fell into one bin. Only the counts in the bins were used, not the individual duration times. If H_0 is true, there are two unknown parameters (ξ and σ, or α and β, or whatever the two parameters of the assumed distribution are).

If H_1 is true, there are $c - 1$ parameters. These parameters are Pr(T falls in bin 1), Pr(T falls in bin 2), etc. There are c bins, but only $c - 1$ parameters, because the probabilities must sum to 1.0. Thus, any $c - 1$ parameters determine the final one.

The big general theorem should say that the degrees of freedom are $(c - 1) - 2$. However, a subtle complication arises. An assumption of the big theorem is that the two unknown parameters of the distribution are estimated using the maximum likelihood estimated based on the bin counts. In this setting, however, it is far easier to estimate those parameters from the raw data. Therefore, as stated in Section 6.6.2.3.2, the degrees of freedom fall somewhere between $c - 3$ and $c - 1$.

To summarize this section, the Pearson chi-squared test has many applications. To avoid confusion, the analyst must clearly specify the null hypothesis being tested and the alternative hypothesis.

7.5.3 Nonparametric Estimators of $\lambda(t)$

The nonparametric estimators of a density in Section 6.6.3 can also be used to estimate a time-dependent event-occurrence rate $\lambda(t)$. In each case, the data consist of a number of times, durations in Section 6.6.3 and event times in the present case. The only difference is the scale: a density integrates to 1.0, and an occurrence rate does not. To use an estimator from Section 6.6.3 in the occurrence-rate setting, multiply the estimate by the total number of observed events. This is the correct scale correction.

The estimators in Section 6.6.3.1.1 showed a problem at 0, estimating a positive density to the left of zero even though a duration time cannot be negative. This problem was corrected by reflecting the data around 0, and initially using a data set that contained both the true data and the mirror images (a value at $-t$ for every value of t). Such a problem also occurs with estimation of $\lambda(t)$. If data are collected in a time period from T_0 to T_1, simple kernel estimates will have this problem at both ends. To correct this problem, reflect the data at each end, so that artificial data have been constructed beyond T_0 on the left and beyond T_1 on the right. Construct the density estimate based on this augmented data set. Then truncate the density — set it to zero outside the observation interval, and multiply it by 3 so that it again integrates to 1.0. Finally, convert it from a density estimate to an estimate of the Poisson intensity by multiplying it by the number of observed events. When interpreting the resulting graphed function, be aware that the estimate will be flat at the two ends, by the way the estimate was constructed. The slope of the line at the two ends cannot be used to draw inferences about whether the Poisson intensity is changing.

This method was applied to the data of Table 7.3 (Example 2.1), unplanned scrams at a new reactor. The normal kernel was used. The 34 event dates were converted to consecutive days, and the standard deviation was 874 days. The formula

$h = 1.06\,\sigma n^{-1/5}$

resulted in h = 457 days. This is only a very preliminary suggested value for two reasons: F was estimated, not known, and the true shape of the intensity function is not close to normal. Because it is wise to undersmooth rather than oversmooth, a value h = 400 days was used.

The estimated Poisson intensity function, $\delta(t)$, is shown in Figure 7.20. The calculations were performed in terms of calendar days, and converted to reactor-critical-years by assuming 0.82 critical years per calendar year. This is the average of the values shown in the right column of Table 7.4.

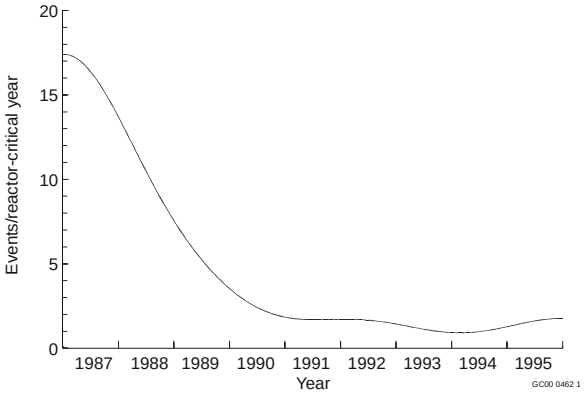


Figure 7.20 Estimate of $\delta(t)$ using standard normal kernel, with bandwidth h = 400.

The curve in Figure 7.20 can be compared with the exponential curve in Figure 7.12. The curve using the kernel estimator follows the ups and downs of the data more closely than does the simple exponential curve. However, the kernel-estimator curve at the beginning of 1987 is only 17.4, substantially below the simple maximum likelihood estimate of 26.8, based on the 1987 data only. Two factors contribute to this. First, the bandwidth of h = 400 days is apparently too wide. The learning period at the plant lasted less than one year, so a smaller bandwidth would better reflect the rapid learning that was taking place. Second, the conversion from calendar time to reactor-critical-years used the average value for nine years, 0.82. In fact, the first calendar year had only 0.71 critical years. Therefore, the estimate during the first year should be about 15% larger than shown.

Figure 7.21 shows the kernel estimator with a smaller bandwidth, h = 200 days. It follows the rapid drop in the scram frequency at the beginning of the plant's history much more closely. It also is more sensitive to small, perhaps random, clusters of events later in the plant's history. The constant conversion rate of 0.82 reactor-critical-years per calendar year has been used, with the same effect as in the previous figure.

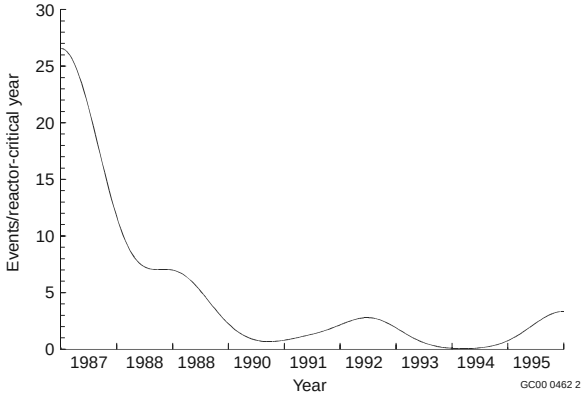


Figure 7.21 Estimate of $\delta(t)$ using standard normal kernel, with bandwidth h = 200.

This example has illustrated both some advantages and some difficulties of nonparametric estimation of a Poisson intensity function. The same advantages and difficulties were seen for nonparametric density estimation. See Section 6.6.3 for more discussion of these issues.

8. PARAMETER ESTIMATION USING DATA FROM DIFFERENT SOURCES

8.1 The Hierarchical Model

This chapter contains the most complex mathematics of the handbook, although most of the formulas can be skipped by those who use preprogrammed software. The formulas are of interest primarily to those who must write the programs. Comments throughout the chapter state which sections can be skipped by most readers.

In this chapter, data come from a number of similar, but not identical, sources. For simplicity, the discussion is in terms of data from a number of nuclear power plants. However, the ideas can be applied much more widely.

The situation is described by a **hierarchical model**, with two levels. The first level models the plants as a family, with the members resembling each other. The second level models the data that are generated at each plant.

To be more specific, suppose that initiating events are to be modeled, so the parameter of interest is λ. Level 1 of the model says that λ varies among the m plants, but only to a limited degree. Thus, the plants are not identical, but they resemble each other. This is modeled by a distribution g that describes the population variability. Before any data are generated, the distribution g is invoked m times, producing values λ_1 through λ_m. These values of λ_i are independently generated, but they all come from the same distribution, g. For each i, λ_i is assigned to plant i. That is Level 1 of the hierarchical model. It is shown on the left side of Figure 8.1.

Level 2 of the model says that, conditional on the λ_i values, the plants independently produce data. Thus, for each i, plant i is observed for time t_i, and it experiences a random number of initiating events, X_i, with X_i having a Poisson($\lambda_i t_i$) distribution. This is shown on the right side of Figure 8.1.

The population-variability distribution g could be a gamma(α, β) distribution, which has the computational advantage of being conjugate to the Poisson distribution. But that is not the only allowed distribution. It could also be a lognormal(μ, σ^2) distribution, or some other distribution.

The data consist of the observation times, t_1 through t_m, which are known and treated as fixed, and the event counts, x_1 through x_m, which are treated as randomly generated. The unknown parameters consist of λ_1 through λ_m, as well as any unknown parameters of g. These parameters of g could be α and β, or μ and σ^2, or some other parameters, depending on the assumed form of g. To emphasize the contrast between the two levels, the parameters of g, such as α and β or μ and σ^2, are sometimes called **hyperparameters**.

When the data instead are failures on demand, the situation is very similar. The population-variability distribution generates parameters p, one for each plant. The distribution g might be a beta distribution, or it might be some nonconjugate distribution, such as (truncated) lognormal or logistic-normal. The unknown parameters consist of the parameters of g, and the parameters p_1 through p_m. The data consist of the counts of failures and demands at each plant, (x_1, n_1) through (x_m, n_m). Examples 8.1 and 8.2 illustrate the two types of data.

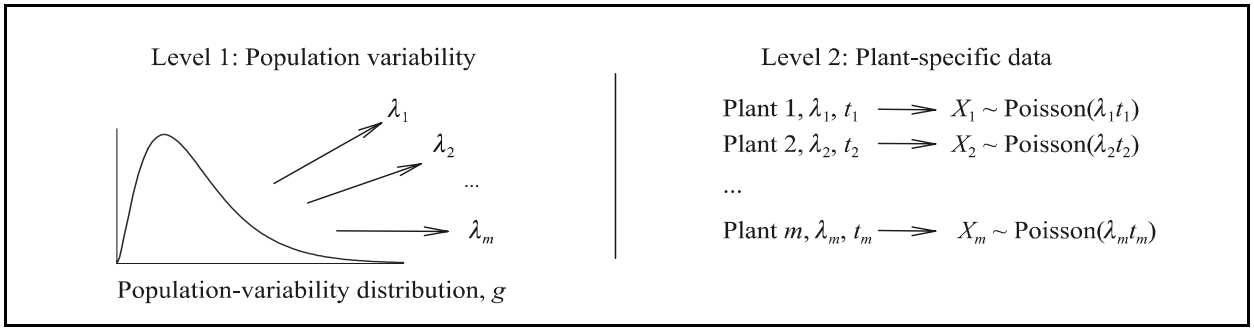


Figure 8.1 Hierarchical model for Poisson data.

Example 8.1 Initiating events from many plants.

The number of unplanned scrams at power, x, and the number of 1000 critical hours, t, are listed below for one calendar year (1984) at 66 plants. The data are from Martz et al. (1999).

Plant	x	t	Plant	x	t	Plant	x	t	Plant	x	t
Arkansas 1	3	6.2500	Farley 2	6	8.3333	Monticello	0	0.8106	San Onofre 2	5	5.2632
Arkansas 2	12	7.6433	Fort Calhoun	1	5.2632	North Anna 1	8	4.7619	San Onofre 3	7	5.0725
Beaver Valley 1	4	6.4516	Ginna	1	6.6667	North Anna 2	4	6.1538	St. Lucie 1	6	5.5556
Big Rock Point	2	6.8966	Grand Gulf	7	2.0896	Oconee 1	3	7.5000	St. Lucie 2	9	7.3770
Brunswick 2	3	2.6549	Haddam Neck	3	6.5217	Oconee 2	0	8.7840	Summer	11	5.5556
Callaway	12	1.5038	Hatch 1	7	5.6452	Oconee 3	4	6.5574	Surry 1	8	5.2980
Calvert Cliffs 1	5	7.5758	Hatch 2	7	3.1111	Oyster Creek	2	1.6949	Surry 2	14	7.4468
Cook 1	3	8.1081	Indian Point 2	4	4.7059	Palisades	1	1.5625	Susquehanna 1	7	6.5421
Cook 2	7	5.3030	Indian Point 3	7	6.9307	Pt. Beach 1	0	6.4201	Susquehanna 2	7	2.1472
Cooper Station	3	6.0000	Kewaunee	4	7.5472	Pt. Beach 2	0	7.5442	Turkey Point 3	8	7.3394
Crystal River 3	2	8.3333	LaSalle 1	9	6.2937	Prairie Island 1	4	8.3333	Turkey Point 4	9	5.0847
Davis-Besse	4	5.5556	LaSalle 2	11	5.4726	Prairie Island 2	0	7.8440	Vermont Yank.	2	7.1429
Diablo Canyon 1	5	1.0846	Maine Yankee	7	6.6667	Quad Cities 1	3	4.7619	Wash. Nucl. 2	23	4.3643
Dresden 2	3	6.5217	McGuire 1	4	6.0606	Quad Cities 2	2	6.8966	Zion 1	6	6.3158
Dresden 3	8	3.8835	McGuire 2	16	6.9869	Robinson 2	0	0.6161	Zion 2	7	6.3063
Duane Arnold	6	6.5934	Millstone 1	0	6.9902	Salem 1	10	2.6738			
Farley 1	2	6.8966	Millstone 2	3	8.5714	Salem 2	10	3.3898			

Example 8.2 Failure to start of AFW motor-driven segments at many plants.

The number of failures to start on unplanned demands for motor-driven segments of the auxiliary feedwater (AFW) system are tabulated for 68 plants, for 1987-1995. Here, x is the number of failures and n is the number of demands. Common-cause failures are excluded. The data are from Poloski et al. (1998, Table E-4).

Plant	x	n	Plant	x	n	Plant	x	n	Plant	x	n
Arkansas 1	0	14	Crystal River 3	1	16	North Anna 2	0	18	Seabrook	0	17
Arkansas 2	0	9	Diablo Canyon 1	0	46	Oconee 1	0	18	Sequoyah 1	0	30
Beaver Valley 1	0	24	Diablo Canyon 2	0	30	Oconee 2	0	18	Sequoyah 2	0	41
Beaver Valley 2	0	43	Farley 1	0	34	Oconee 3	0	12	South Texas 1	0	69
Braidwood 1	0	13	Farley 2	0	54	Palisades	0	13	South Texas 2	0	87
Braidwood 2	0	24	Fort Calhoun	0	5	Palo Verde 1	0	7	St. Lucie 1	0	35
Byron 1	0	11	Ginna	0	28	Palo Verde 2	0	12	St. Lucie 2	0	21
Byron 2	0	26	Harris	0	98	Palo Verde 3	0	9	Summer	0	24
Callaway	0	57	Indian Point 2	1	24	Point Beach 1	0	8	Surry 1	0	26
Calvert Cliffs 1	0	12	Indian Point 3	2	32	Point Beach 2	0	16	Surry 2	0	32
Calvert Cliffs 2	0	15	Kewaunee	0	26	Prairie Island 1	0	3	Three Mile Island 1	0	6
Catawba 1	0	41	Maine Yankee	0	23	Prairie Island 2	0	7	Vogtle 1	0	103
Catawba 2	0	89	McGuire 1	0	45	Robinson 2	1	28	Vogtle 2	0	45
Comanche Pk 1	0	66	McGuire 2	0	44	Salem 1	0	24	Waterford 3	0	38
Comanche Pk 2	0	14	Millstone 2	1	11	Salem 2	0	32	Wolf Creek	0	51
Cook 1	0	18	Millstone 3	0	54	San Onofre 2	0	13	Zion 1	0	13
Cook 2	0	36	North Anna 1	0	20	San Onofre 3	0	17	Zion 2	0	8

In Example 8.1, most of the plants experience at least one initiating event, and the total number of events is 361. Thus, the data set is large, and the methods given below perform well. Example 8.2, on the other hand, is a small data set. That is, most of the plants experienced no failures, and the total number of failures is only 6. This example was deliberately chosen for this handbook to illustrate problems that can occur with sparse data.

Two methods are given in Sections 8.2 and 8.3 for analyzing data by means of a hierarchical model. The results of each analysis include both an estimate of the population-variability distribution, g, and estimates of all the plant-specific parameters, θ_i or p_i.

8.2 The Parametric Empirical Bayes Method

8.2.1 General Approach

In spite of the name, this is not a truly Bayesian method. Instead, it is a kind of hybrid, involving a non-Bayesian step followed by a Bayesian step.

Step 1. Look at the variability in the data from the plants, and estimate g. That is, based on the data from all the plants, estimate the parameters of g by maximum likelihood, and obtain the resulting estimate of the distribution. Call the estimate $\hat{g}$.

Step 2. Now treat $\hat{g}$ as a prior distribution. Perform the usual Bayesian update, with the prior distribution $\hat{g}$ and data from a single plant, to get the posterior distribution for the plant-specific parameter, θ_i or p_i.

Thus, the method yields both an estimate of the population variability and plant-specific estimates at each plant.

The method as just explained underestimates the uncertainty in the answers, because it treats $\hat{g}$ as if it were equal to the true distribution g. Therefore, the best implementations of the empirical Bayes method add a final adjustment to Step 2, which makes the plant-specific posterior distributions somewhat more diffuse. This largely accounts for the inaccuracy in equating $\hat{g}$ to g.

The name "empirical Bayes" follows from the two steps. The plant-specific estimates are found in a Bayesian way, by updating a prior distribution with plant-specific data. However, the prior is not based on prior data or on prior belief, but instead is based on the existing data — the prior is determined empirically.

Step 1 can be carried out in a simple way only if the distributions have convenient forms. Thus, parametric empirical Bayes estimation assumes that g is conjugate to the distribution of the data at any plant. That is, g is a gamma distribution when the data are initiating events or other data from a Poisson process, and g is a beta distribution when the data are failures on demand. This is a limitation to the method. One reason for introducing the hierarchical Bayes method, in Section 8.3, is to overcome this limitation.

Some people might object that the method double counts the data. It uses the data to decide on the prior distribution, and then it uses the same data again to update the prior to obtain the plant-specific estimates. There are two responses to this: (1) The objection is not important in practice, unless the number of plants in the study is very small, or if a small number of plants dominate the data. If no single plant contributes much to the estimate of g, then there is very little double counting that influences the final estimate for that plant. (2) The hierarchical Bayes method, given in Section 8.3, will avoid this difficulty entirely.

For failures on demand, Martz et al. (1996) give a tutorial on the empirical Bayes method, illustrated with nuclear power plant (NPP) data. Siu and Kelly (1998) also explain the method as part of their tutorial article. Carlin and Louis (2000) give a full treatment, including worked-out examples.

8.2.2 MLE Equations for the Gamma-Poisson Model

Readers who do not need to know the equations can skip directly to Section 8.2.5.

The gamma-Poisson model is used for initiating events. The data at plant i consist of a count of events, x_i, in time t_i. Conditional on the plant-specific parameter θ_i, it is assumed that x_i was generated from a Poisson($\theta_i t_i$) distribution. However θ_i was generated from the distribution g, which is assumed to be gamma(α, β). The equations for the maximum likelihood estimates (MLEs) of α and β are now given.

The *conditional distribution* of X, conditional on θ, is Poisson. However, the *unconditional distribution* of X, when θ might be any value generated by the population-variability distribution g, is more complicated. It can be shown that the unconditional distribution equals the distribution conditional on θ, averaged over the possible values of θ. In equation form, this is

$$\Pr(X = x|\alpha,\beta) = \int \Pr(X = x|\lambda) g(\lambda|\alpha,\beta) d\lambda .$$

Substituting the formulas for the Poisson and gamma distributions, it can be shown that this equals

$$\Pr(X = x|\alpha,\beta) = \frac{\Gamma(\alpha+x)}{n!\Gamma(\alpha)}(t/\beta)^x(1+t/\beta)^{-(\alpha+x)} . \quad (8.1)$$

As mentioned in Section 6.2.3.5, Equation 6.8, this is the gamma-Poisson distribution, also often called the negative binomial distribution. This distribution is not conditional on θ. Therefore, θ does not appear in this expression. Instead, Equation 8.1 gives the probability of seeing x events in time t at a plant with a randomly assigned θ.

To find the MLEs of α and β, write the product of terms of the form of Expression 8.1, using values (x_i, t_i) for i from 1 to m. That product is the joint unconditional likelihood of the data. Take the logarithm of that expression and maximize it. There are several ways to accomplish this.

One approach is to maximize it numerically as a function of two variables, using some version of Newton's method. This is the approach of Siu and Kelly (1998). The derivatives of the log-likelihood, used in performing the maximization, are given below, as stated by Engelhardt (1994). Here $\ln L$ denotes the logarithm of the likelihood.

$$\frac{\partial}{\partial\alpha}\ln L = \sum_{i=1}^{m}[\psi(\alpha + x_i) - \psi(\alpha) - \ln(1 + t_i / \beta)]$$

$$\frac{\partial}{\partial\beta}\ln L = -\frac{1}{\beta}\sum_{i=1}^{m}\left[x_i - \frac{\alpha + x_i}{\beta + t_i}t_i\right] .$$

The function ψ is the **digamma function**, $\psi(u) = (d/du)\ln \Gamma(u)$. It is built into many computer packages. Because x_i is an integer, the expression involving ψ can be rewritten as

$$\psi(\alpha + x_i) - \psi(\alpha) = \sum_{j=1}^{x_i}\frac{1}{\alpha + j - 1} . \qquad (8.2)$$

A second approach reduces the problem to solving one equation, as follows. At the maximum of the log-likelihood, the two derivatives are equal to zero. Therefore, do the following:

1. Set the two derivatives equal to zero. The solutions, to be found, are $\hat{\alpha}$ and $\hat{\beta}$.
2. Solve the second equation for $\hat{\alpha}$, as a function of $\hat{\beta}$.
3. Substitute this expression into the first equation.
4. Solve the resulting equation numerically for $\hat{\beta}$.
5. Calculate $\hat{\alpha}$ from the numerical value of $\hat{\beta}$.

The necessary equations to carry out these steps are the following. The equation for $\hat{\alpha}$, as a function of $\hat{\beta}$, is

$$\hat{\alpha} = \left(\sum_{i=1}^{m}\frac{x_i\hat{\beta}}{\hat{\beta} + t_i}\right) \Big/ \left(\sum_{i=1}^{m}\frac{t_i}{\hat{\beta} + t_i}\right) . \qquad (8.3)$$

Substitute Equation 8.3 into

$$\sum_{i=1}^{m}[\psi(\hat{\alpha} + x_i) - \psi(\hat{\alpha}) - \ln(1 + t_i / \hat{\beta})] = 0$$

and solve that equation numerically for $\hat{\beta}$. Having obtained the numerical value of $\hat{\beta}$, find $\hat{\alpha}$ from Equation 8.3.

Sometimes the equations do not have a solution. If the plants do not appear to differ much — for example, the naive plant-specific estimates x_i/t_i are all similar — the maximum likelihood estimate of g may be degenerate, concentrated at a single point. That says that the plants appear to have a single common θ. Engelhardt (1994) recommends aborting the estimation process, not trying to fit a model, if the estimate of β becomes greater than Et_i during the iterations. The population-variability distribution g would be gamma(α, β), with the second parameter greater than Et_i. But simply pooling the data (and using a Jeffreys prior) would result in a gamma(Ex_i + ½, Et_i) posterior distribution. Thus the empirical Bayes distribution would produce a between-plant distribution that is more concentrated (larger second parameter) than the distribution when the plant data are pooled. This is not the intent of the hierarchical model.

8.2.3 MLE Equations for the Beta-Binomial Model

The beta-binomial model is used for failures on demand. The data at plant i consist of a count of failures, x_i, and demands, n_i. Conditional on the plant-specific parameter p_i, it is assumed that x_i was generated from a binomial(n_i, p_i) distribution. However, p_i was generated from the distribution g, which is assumed to be beta(α, β). The equations for the MLEs of α and β are now given.

The *conditional distribution* of X, conditional on p, is binomial. However, the *unconditional distribution* of X, when p might be any value generated by the population-variability distribution g, equals the distribution conditional on p, averaged over the possible values of p. That is,

$$\Pr(X = x|\alpha,\beta) = \int \Pr(X = x|p)g(p|\alpha,\beta)dp .$$

Substituting the formulas for the binomial and beta distributions, and using some standard identities relating the beta function and the gamma function, it can be shown that this equals

$$\Pr(X = x) = \frac{n!}{x!(n-x)!}\frac{\Gamma(\alpha+x)}{\Gamma(\alpha)}\frac{\Gamma(\beta+n-x)}{\Gamma(\beta)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)} .$$

This is called the beta-binomial distribution. If both x and n are integers, this can be rewritten without the gamma function, as follows:

$$\Pr(X = x) = \frac{n!}{x!(n-x)!}\prod_{j=0}^{x-1}(\alpha+j)\prod_{j=0}^{n-x-1}(\beta+j)\Big/\prod_{j=0}^{n-1}(\alpha+\beta+j) .$$

As just stated, this is valid if x and n are integers. Are they not always integers? No, they are not, if the data set only gives an *estimate* of the number of demands, which is not necessarily an integer. In that case, the expression with the gamma function is the only one that can be used.

The likelihood is the product of terms of one of these forms, containing values (x_i, n_i) for i = 1 to m. To find the MLE, take the logarithm of the likelihood and maximize it.

The maximization can be done in a variety of ways. One approach, following Atwood (1994), does not deal with α and β directly. Instead, it reparameterizes, working with

$\delta = \alpha + \beta$ and
$\mu = \alpha/\delta$.

The intuitive reason for this reparameterization is that μ is the mean of the binomial distribution, and in most models the mean is one of the easiest things to estimate. The letter δ was chosen as a mnemonic for "dispersion," because the variance of the binomial distribution is $\mu(1-\mu)/(\delta+1)$. Thus, δ is related directly to the variance. Recall that in Section 6.3.2.2.2, the prior and posterior values of $\alpha+\beta$, which we are calling δ here, were interpreted as the prior and posterior number of demands.

After working with μ and δ, and finding the MLEs of these parameters, we will translate back to find the MLEs of α and β, using the equations:

$\alpha = \mu\delta$
$\beta = (1-\mu)\delta$.

The MLE is found by setting the derivatives with respect to μ and δ to zero. After some manipulation, the equations can be expressed as

$$\sum_{j=1}^{m}\{\psi(\mu\delta+x_i)-\psi(\mu\delta)\} = \sum_{i=1}^{m}\{\psi((1-\mu)\delta+n_i-x_i)-\psi((1-\mu)\delta)\} \quad (8.4a)$$

$$\sum_{j=1}^{m}\{\psi((1-\mu)\delta+n_i-x_i)-\psi((1-\mu)\delta)\} = \sum_{j=1}^{m}\{\psi(\delta+n_i)-\psi(\delta)\} . \quad (8.4b)$$

Here Ψ is the digamma function, the derivative of $\ln\Gamma$, just as in Section 8.2.2. If x_i and n_i are integers for all i, Equation 8.2 can be used to rewrite Equation 8.4 as

$$\sum_{i=1}^{m}\left[\sum_{j=0}^{x_i-1}\frac{1}{\mu\delta+j}\right] = \sum_{i=1}^{m}\left[\sum_{j=0}^{n_i-x_i-1}\frac{1}{(1-\mu)\delta+j}\right] \quad (8.5a)$$

and

$$\sum_{i=1}^{m}\left[\sum_{j=0}^{n_i-x_i-1}\frac{1}{(1-\mu)\delta+j}\right] = \sum_{i=1}^{m}\left[\sum_{j=0}^{n_i-1}\frac{1}{\delta+j}\right] . \quad (8.5b)$$

The Equations 8.4 or 8.5 must be solved for μ and δ. One method, suggested by Atwood (1994) is to begin with a trial value of δ. Solve Equation 8.4a or 8.5a numerically for μ. This typically needs only a few iterations. Substitute this value into Equation 8.4b or 8.5b, and solve the resulting equation for δ. Continue alternating between the two equations until the estimates converge.

The estimates do not always converge. If the plants have very similar data, the maximum likelihood estimate of g may be concentrated at a single point, degen-

erate. This would say that the plants all have the same *p*. Atwood (1994) recommends aborting the iterations if the value of * becomes greater than En_i. Allowing * to be greater than En_i would produce a population-variability distribution that is more concentrated than the distribution corresponding to simply pooling the data.

8.2.4 Adjustment for Uncertainty in the Estimate of *g*

As mentioned above, the method as presented so far underestimates the uncertainty in the final answers, because it does not account for the uncertainty in $\hat{g}$. Kass and Steffey (1989) present a correction to the final estimates, to approximately account for this uncertainty. The plant-specific posterior means are unchanged, but the posterior variances are increased somewhat. Kass and Steffey (1998) state that the adjustment is very important if there are few data subsets (plants, in the present discussion) and many observations (initiating events or demands). Conversely, the adjustment is unimportant when there are many data subsets and few observations. No harm is done by automatically applying the adjustment in every case. The formulas are given here.

8.2.4.1 Equations for the Gamma-Poisson Case

With the gamma-Poisson distribution, it is computationally advantageous to reparameterize in terms of : = "/$ and ". Denote the maximum likelihood estimators for the hyperparameters : and " by $\hat{\mu}$ and $\hat{\alpha}$. It turns out that these estimators are asymptotically uncorrelated, causing certain terms in the formulas to be zero.

The method as given in Section 8.2.2 finds the estimates $\hat{\alpha}$ and $\hat{\beta}$, which can be reparameterized as $\hat{\mu} = \hat{\alpha} / \hat{\beta}$ and $\hat{\alpha}$. These are the estimated parameters of the gamma prior distribution *g*. The method then updates the estimated prior by plant-specific data. The posterior distribution of θ_i is also a gamma distribution, with posterior mean

$$E_{\text{post}}(\lambda_i) = \frac{\hat{\alpha} + x_i}{\hat{\alpha} / \hat{\mu} + t_i} \tag{8.6}$$

and posterior variance

$$\text{var}_{\text{post}}(\lambda_i) = \frac{\hat{\alpha} + x_i}{\left(\hat{\alpha} / \hat{\mu} + t_i\right)^2} \ .$$

The Kass-Steffey adjustment increases the variance to

$$\text{var}_{\text{post}}(\lambda_i) + \left[\frac{\partial E_{\text{post}}(\lambda_i)}{\partial \hat{\mu}}\right]^2 \text{var}(\hat{\mu}) + \left[\frac{\partial E_{\text{post}}(\lambda_i)}{\partial \hat{\alpha}}\right]^2 \text{var}(\hat{\alpha}) \ . \tag{8.7}$$

A covariance term would normally also be present, but this term is zero when the parameterization is in terms of : and ".

We now develop the formulas that must be substituted into Equation 8.7. From Equation 8.6, the derivatives of $E_{\text{post}}(\lambda_i)$ are

$$\frac{\partial E_{\text{post}}(\lambda_i)}{\partial \hat{\mu}} = -\frac{\hat{\alpha}(\hat{\alpha} + x_i)}{(\hat{\alpha} + \hat{\mu} t_i)^2} \quad \text{and}$$

$$\frac{\partial E_{\text{post}}(\lambda_i)}{\partial \hat{\alpha}} = -\frac{\hat{\mu}(x_i - \hat{\mu} t_i)}{(\hat{\alpha} + x_i \hat{\mu})^2} \ .$$

From the asymptotic theory of maximum likelihood estimation, the variances are found as follows. The information matrix, *J*, is the negative of the matrix of second derivatives of the log-likelihood:

$$J = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} = -E\begin{bmatrix} \dfrac{\partial^2 L}{\partial \mu^2} & \dfrac{\partial^2 L}{\partial \mu \partial \alpha} \\ \dfrac{\partial^2 L}{\partial \mu \partial \alpha} & \dfrac{\partial^2 L}{\partial \alpha^2} \end{bmatrix} \tag{8.8}$$

evaluated at $\hat{\mu}$ and $\hat{\alpha}$. The inverse of this matrix is asymptotically equal to the variance-covariance matrix:

$$\begin{bmatrix} \text{var}(\hat{\mu}) & \text{cov}(\hat{\mu}, \hat{\alpha}) \\ \text{cov}(\hat{\mu}, \hat{\alpha}) & \text{var}(\hat{\alpha}) \end{bmatrix} .$$

When this is carried out, we have

$$J_{11} = \frac{\hat{\alpha}}{\hat{\mu}} \sum_{i=1}^{m} \frac{t_i}{\hat{\alpha} + \hat{\mu} t_i}$$

$$J_{22} = \sum_{i=1}^{m} \left[\psi'(\hat{\alpha}) - \psi'(\hat{\alpha} + x_i)\right] - \frac{\hat{\mu}}{\hat{\alpha}} \sum_{i=1}^{m} \frac{t_i}{\hat{\alpha} + \hat{\mu} t_i}$$

$$J_{12} = J_{21} = 0 \ .$$

If all the x_i values are integers, the difference of ψ terms can be rewritten using Equation 8.2, and the difference of derivatives ψ' can be written algebraically, avoiding use of a special function.

Because the off-diagonal elements are zero, the inverse consists of the inverses of the diagonal terms, and it follows that

$$\text{var}(\hat{\mu}) = 1/J_{11} \quad \text{and}$$
$$\text{var}(\hat{\alpha}) = 1/J_{22} \ .$$

The final step of the empirical Bayes method is to substitute the expressions just found into the Kass-Steffey adjustment for the posterior variance, Equation 8.7. Then approximate the posterior distribution by a gamma distribution having the original posterior mean and the adjusted posterior variance. An example will be given below.

8.2.4.2 Equations for the Beta-Binomial Case

As in Section 8.2.3, we parameterize in terms of $:$ = $''/(\ '' + \$)$ and $*$ = $\ '' + \$$. Denote the maximum likelihood estimators by $\hat{\mu}$ and $\hat{\delta}$. Although these estimators are asymptotically not exactly uncorrelated, as was the case for the gamma-Poisson model, they are nearly uncorrelated. The equations are given by Atwood (1995).

The method as given in Section 8.2.3 finds the estimates $\hat{\mu}$ and $\hat{\delta}$, the estimated parameters of the beta prior distribution *g*. The method then updates the estimated prior by plant-specific data. The posterior distribution of p_i is also a beta distribution, with posterior mean

$$E_{\text{post}}(p_i) = \frac{\hat{\mu}\hat{\delta} + x_i}{\hat{\delta} + n_i} \tag{8.9}$$

and posterior variance

$$\text{var}_{\text{post}}(p_i) = E_{\text{post}}(p_i)[1 - E_{\text{post}}(p_i)] / (\hat{\delta} + n_i + 1) \ .$$

The Kass-Steffey adjustment increases the variance to

$$\text{var}_{\text{post}}(p_i) + \left[\frac{\partial E_{\text{post}}(p_i)}{\partial \hat{\mu}}\right]^2 \text{var}(\hat{\mu}) + \left[\frac{\partial E_{\text{post}}(p_i)}{\partial \hat{\delta}}\right]^2 \text{var}(\hat{\delta}) + \left[\frac{\partial E_{\text{post}}(p_i)}{\partial \hat{\mu}}\right]\left[\frac{\partial E_{\text{post}}(p_i)}{\partial \hat{\delta}}\right] \text{cov}(\hat{\mu}, \hat{\delta}) \ . \tag{8.10}$$

From Equation 8.9, the two derivatives of $E_{\text{post}}(p_i)$ are

$$\frac{\partial E_{\text{post}}(p_i)}{\partial \hat{\mu}} = \frac{\hat{\delta}}{\hat{\delta} + n_i}$$

$$\frac{\partial E_{\text{post}}(p_i)}{\partial \hat{\delta}} = -\frac{x_i - n_i\hat{\mu}}{(\hat{\delta} + n_i)^2} \ .$$

The variances and covariance are found from inverting the matrix in Equation 8.8, with $*$ used now instead of $''$. The terms can be found as follows. Define

$$S_1 = \sum_{i=1}^{m}\left[\psi'(\hat{\mu}\hat{\delta}) - \psi'(\hat{\mu}\hat{\delta} + x_i)\right]$$

$$S_2 = \sum_{i=1}^{m}\left[\psi'((1-\hat{\mu})\hat{\delta}) - \psi'((1-\hat{\mu})\hat{\delta} + n_i - x_i)\right]$$

$$S_3 = \sum_{i=1}^{m}\left[\psi'(\hat{\delta}) - \psi'(\hat{\delta} + n_i)\right] \ .$$

Then the information matrix is given by

$$J_{11} = \hat{\delta}^2(S_1 + S_2)$$
$$J_{22} = \hat{\mu}^2 S_1 + (1-\hat{\mu})^2 S_2 - S_3$$
$$J_{12} = J_{21} = \hat{\mu}\,\hat{\delta} S_1 - (1-\hat{\mu})\hat{\delta} S_2 \ .$$

The variances and covariance follow from standard formulas for inverting a 2×2 matrix. Define the determinant

$$D = J_{11}J_{22} \ ! \ (J_{12})^2 \ .$$

Then we have

$$\text{var}(\hat{\mu}) = J_{22} / D$$
$$\text{var}(\hat{\delta}) = J_{11} / D$$
$$\text{cov}(\hat{\mu}, \hat{\delta}) = -J_{12} / D \ .$$

To complete the Kass-Steffey adjustment, substitute the above expressions into the equation for the adjusted variance, Equation 8.10. Then find the beta distribution having the posterior mean and the adjusted posterior variance. Use this as the approximate posterior distribution for p_i.

8.2.5 Application to Examples

The parametric empirical Bayes method is now illustrated with Examples 8.1 and 8.2. First, a chi-squared test will be performed, to test whether the plants can be

pooled. In each example, the difference between plants will be found to be either statistically significant or very close to statistically significant. Then plant-specific confidence intervals will be found, each based only on the data for a single plant. Then the empirical Bayes method will be used, and the resulting 90% credible intervals will be shown, based on the plant-specific posterior distributions, using the Kass-Steffey adjustment. The plant-specific intervals resulting from the empirical Bayes analysis will be compared to the (less sophisticated) plant-specific confidence intervals.

8.2.5.1 Example 8.1, Initiating Events

To test poolability of the plants in Example 8.1, the Pearson chi-squared test was performed, as presented in Section 6.2.3.1.2. The test statistic X^2 was equal to 378.5. Because there were 66 plants, the value of X^2 should be compared to a chi-squared distribution with 65 degrees of freedom. The value of 378.5 is very far out in the tail of the chi-squared distribution, off the table. Thus, the evidence is extremely strong, beyond question, that the plants do not all have the same δ.

To show this graphically, 90% confidence intervals for δ were plotted, with each confidence interval based on the data from a single plant. These are shown in Figure 8.2. Because t has been written in terms of 1,000 critical hours, the units of δ are events per thousand critical hours. The order of the plants is not alphabetical, but instead is by decreasing estimate of δ. Because the example has so many plants, only the plants with the 10 highest and 10 lowest estimates are individually identified in the figure.

A 90% confidence interval is plotted at the top of the plot for the pooled industry data. Of course the interval is too short to be realistic, because pooling of the data is completely unjustified in this example. In fact, the interval is too short even to be visible. Nevertheless, the overall pooled mean is a useful reference value for comparison with the individual plant results. For this reason, a vertical dashed line is drawn through the pooled mean.

Because the plant-specific estimates differ so greatly, the figure uses a logarithmic scale. This means that some of the point estimates, those with zero values, cannot be plotted.

Figure 8.3 is based on the empirical Bayes method. For each plant, the mean and 90% credible interval are shown, based on the posterior distribution and the Kass-Steffey adjustment. The mean and 90% interval for the industry are also plotted, and a vertical dashed line is drawn through the mean.

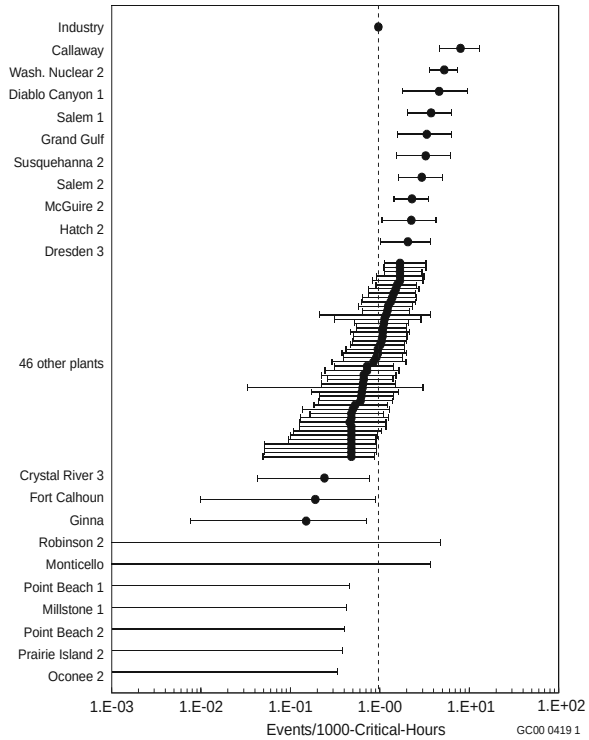


Figure 8.2 Plant-specific MLEs and 90% confidence intervals for δ.

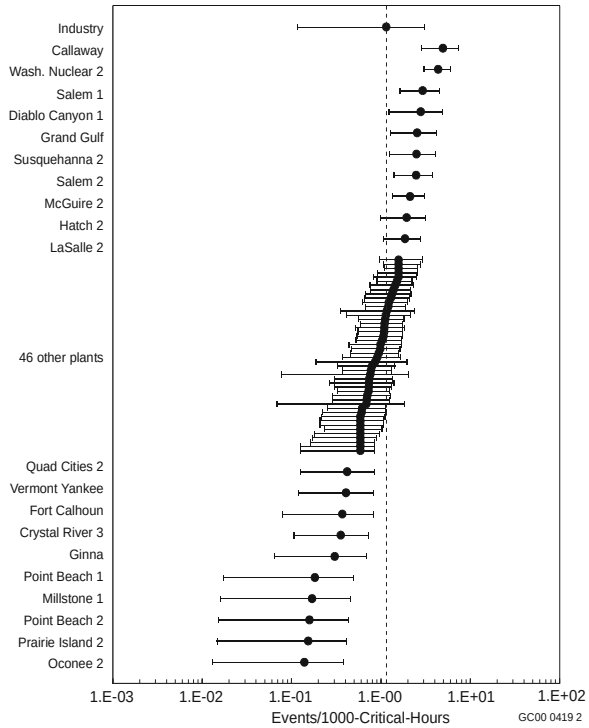


Figure 8.3 Plant-specific posterior means and 90% credible intervals for δ.

The most striking feature seen by comparing the two figures is that the empirical Bayes estimates vary less from each other than do the MLEs. Of course, if a plant has no events, the lower confidence limit is zero, and any Bayesian method will give a non-zero lower limit. Such a difference appears enormous when plotted on a logarithmic scale. However, the effect is seen not only at the bottom of Figures 8.2 and 8.3 but also at the top: the largest plant-specific posterior means are closer to the industry average than are the corresponding MLEs. Indeed, just as was seen for Bayes methods in general, the empirical Bayes method gives posterior means that are between the MLEs and the industry (i.e., the prior) mean.

Those who wish to make some detailed comparisons can find a few numerical values listed in Tables 8.1 through 8.3.

Table 8.1 Portion of frequentist analysis results for Example 8.1.

Plant	x, t	MLE and 90% confidence interval[a]
Industry	361, 374.229	(0.883, 0.965, 1.05)
Callaway	12, 1.5038	(4.60, 7.98, 12.9)
Wash. Nuc. 2	23, 4.3643	(3.60, 5.27, 7.47)
Diablo Can. 1	5, 1.0846	(1.82, 4.61, 9.69)
Salem 1	10, 2.6738	(2.03, 3.74, 6.34)
Grand Gulf	7, 2.0896	(1.57, 3.35, 6.29)
Pt. Beach 2	0, 7.5442	(0.0, 0.0, 0.397)
Prairie Isl. 2	0, 7.8440	(0.0, 0.0, 0.382)
Oconee 2	0, 8.7840	(0.0, 0.0, 0.341)

[a] Format is (lower bound, MLE, upper bound).

The order of the plants is not exactly the same in Figures 8.2 and 8.3. The reason is that estimates for different plants are pulled toward the industry mean by different amounts. This can cause some rearrangement of the ranking of the plants. For example, Salem 1 and Diablo Canyon 1 appear in reverse order in the two figures (and in Tables 8.1 and 8.3). The reason is that Diablo Canyon 1 has about half as much data (5 events in 1,085 hours) as Salem 1 (10 events in 2,674 hours). Therefore Diablo Canyon 1 is pulled more toward the industry mean.

We notice also, by comparing Tables 8.2 and 8.3, that the Kass-Steffey adjustment is very small in this example. The data set is so large that g can be estimated quite well. Any error in equating the estimate to the true distribution is minor, as reflected in the small effect of the Kass-Steffey adjustment.

Table 8.2 Portion of empirical Bayes analysis, without Kass-Steffey adjustment.

Plant	Gamma parameters, α, β	Posterior mean and 90% credible interval[a]
Industry	1.39, 1.211	(0.118, 1.15, 3.07)
Callaway	13.39, 2.715	(2.94, 4.93, 7.34)
Wash. Nuc. 2	24.39, 5.575	(3.03, 4.37, 5.93)
Salem 1	11.39, 3.885	(1.66, 2.93, 4.49)
Diablo Can. 1	6.39, 2.296	(1.25, 2.78, 4.81)
Grand Gulf	8.39, 3.301	(1.29, 2.54, 4.14)
Pt. Beach 2	1.39, 8.755	(0.0164, 0.159, 0.424)
Prairie Isl. 2	1.39, 9.055	(0.0158, 0.154, 0.410)
Oconee 2	1.39, 9.995	(0.0143, 0.139, 0.372)

[a] Format is (lower bound, mean, upper bound).

Table 8.3 Portion of empirical Bayes analysis, with Kass-Steffey adjustment.

Plant	Gamma parameters, α, β	Posterior mean and 90% credible interval[a]
Industry	1.39, 1.211	(0.118, 1.15, 3.07)
Callaway	12.13, 2.460	(2.86, 4.93, 7.47)
Wash. Nuc. 2	23.40, 5.348	(3.00, 4.37, 5.96)
Salem 1	11.03, 3.762	(1.65, 2.93, 4.52)
Diablo Can. 1	6.08, 2.185	(1.22, 2.78, 4.86)
Grand Gulf	8.15, 3.204	(1.27, 2.54, 4.16)
Pt. Beach 2	1.33, 8.382	(0.0151, 0.159, 0.431)
Prairie Isl. 2	1.33, 8.665	(0.0146, 0.154, 0.417)
Oconee 2	1.33, 9.554	(0.0132, 0.139, 0.378)

[a] Format is (lower bound, mean, upper bound).

An empirical Bayes estimator is sometimes called a **shrinkage estimator**, or a **shrinker**, because the method pulls all the MLEs in towards the industry mean. The intuitive justification for such shrinkage is the recognition that extreme data are produced by a combination of extreme parameters and luck. Thus, the plant with the highest observed frequency appears so extreme because of a combination of large θ and some bad luck. Likewise, the plant with the best performance, Oconee 2, which ran for 366 days straight without a single scram, can attribute its perfect performance to a combination of low θ and good luck. The empirical Bayes method tries to remove the effect of luck when estimating the θ values.

As always when performing a statistical analysis, one should try to combine statistical calculations with engineering understanding. It is known that newly licensed plants sometimes experience more initiating events than they do after acquiring more experience. This was mentioned in the discussion of Example 2.1, and it is seen again here.

Of the 66 plants, 9 did not have their commercial starts until 1984 or later. These 9 young plants are all among the 19 with the highest event frequencies. For example, consider the two plants with the highest estimated frequencies, based on the 1984 data. Both of these plants had their commercial starts in December 1984.

The hierarchical model is intended for plants that are nominally identical. The variability among the plants is unexplained, and modeled as random. An important assumption is that each plant is assigned a θ from the same distribution, g. As a result, each plant is as likely as any other to have a large θ or a small θ. The parameters θ_i are called **exchangeable** if any θ_i is as likely as any other to correspond to a particular plant. As discussed by Gelman et al. (1995, Section 5.2), when we know nothing about the plants, exchangeability is a reasonable assumption. When we know the ages of the plants, however, exchangeability is no longer reasonable. The most immature plants are expected to have larger values of θ.

Thus, the analysis of Example 8.1 really should be modified. One way would be to separate the plants into two groups, mature and immature, and perform an empirical Bayes analysis on each group. A more sophisticated way would be to try to model the age of the plant as a continuous explanatory variable. Then the otherwise random θ_i would be multiplied by some function of the age of plant i, a large factor for immature plants and a smaller factor for mature plants. Such models are beyond the scope of this handbook, however.

8.2.5.2 Example 8.2, AFW Segment Failures to Start

This example has 68 plants, with 6 failures in 1993 demands.

Poloski et al. (1998) perform a chi-squared test to see if p is the same at all plants. This test is explained in Section 6.3.3.1.2. The test statistic X^2 equals 113.1. Because there are 68 plants, the degrees of freedom is 67. The reported p-value is 0.0004, meaning that 113.1 is the 99.96th percentile of the chi-squared distribution with 67 degrees of freedom. However, the data set has so few failures that the chi-squared distribution is not a good approximation for the distribution of X^2. The expected number of failures at a plant with 30 demands (a typical number of demands) is 6×30/1993 = 0.09. This is much less than the recommended minimum of 0.5. Therefore, the calculated p-value is quite suspect.

Poloski et al. (1998) chose to model between-plant differences with a hierarchical model, partly because of the above calculated p-value, and partly on the grounds that modeling possible differences between plants is more conservative (reflecting more variability) than simply pooling the data.

The empirical Bayes estimate of the population-variability distribution, g, is a beta (0.137, 36.34) distribution. The mean of this distribution is 3.77E!3. The 5th and 95th percentiles are 5.99E!12 and 2.12E!2. Note, the first parameter of the beta distribution is very small, well below 0.5. As a result, the 5th percentile is unrealistically small, implying less than one failure in one hundred billion demands. This unrealistic lower bound carries over to the posterior distribution of all plants that have zero failures. Figures 8.4 and 8.5 are the analogues of Figures 8.2 and 8.3.

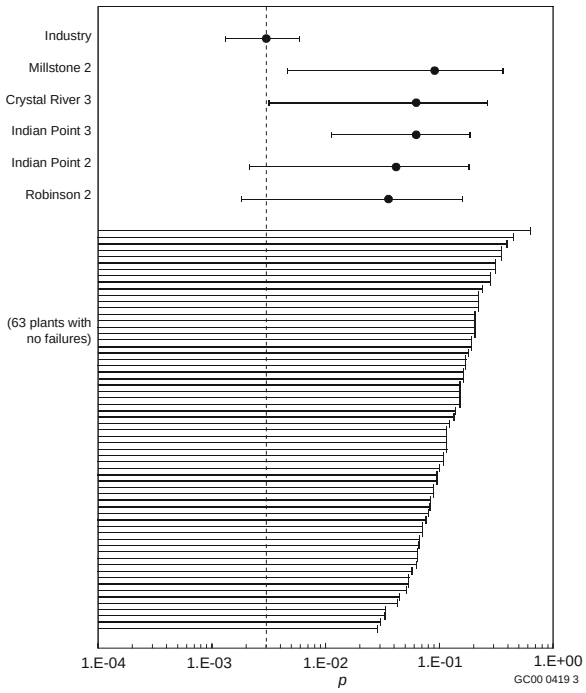


Figure 8.4 Plant-specific MLEs and 90% confidence intervals for p.

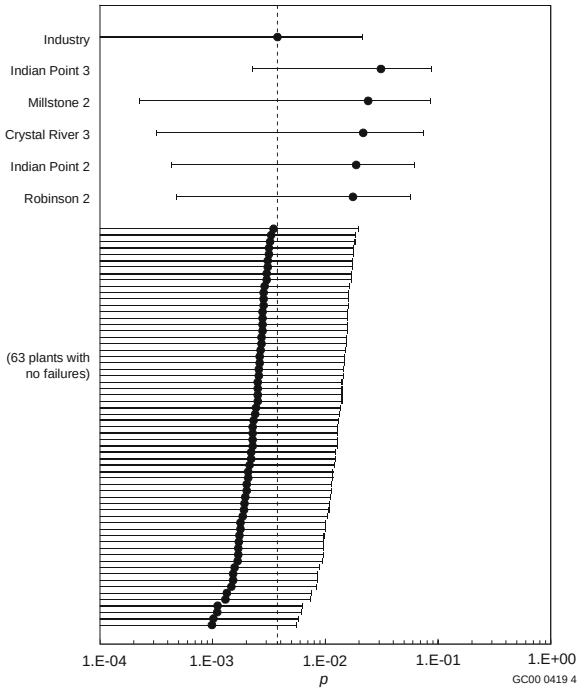


Figure 8.5 Plant-specific posterior means and 90% credible intervals for *p*.

The first figure shows plant-specific MLEs and 90% confidence intervals, while the second shows the results of the empirical Bayes analysis, posterior means and 90% credible intervals. Only the five plants that had failures are individually identified in the figures.

Some numerical details are given in Tables 8.4 through 8.6.

Table 8.4 Portion of frequentist analysis results for Example 8.2.

Plant	*x*, *d*	MLE and 90% conf. int. [a]
Industry	6, 1993	(1.31E!3, 3.01E!3, 5.93E!3)
Millstone 2	1, 11	(4.65E!3, 9.09E!2, 3.64E!1)
Crystal River 3	1, 16	(3.20E!3, 6.25E!2, 2.64E!1)
Indian Point 3	2, 32	(1.12E!2. 6.25E!2, 1.84E!1)
Indian Point 2	1, 24	(2.13E!3, 4.17E!2, 1.83E!1)
Robinson 2	1, 28	(1.83E!3, 3.57E!2, 1.59E!1)
Prairie Island 1	0, 3	(0.0, 0.0, 3.32E!1)
Vogtle 1	0, 103	(0.0, 0.0, 2.87E!2)

[a] Format is (lower bound, MLE, upper bound).

Just as with Example 8.1, the empirical Bayes method pulls the plant-specific MLEs toward the industry mean. This is seen in both the figures and the tables. Also, the Kass-Steffey adjustment increases the width of the plant-specific intervals by a noticeable amount, for example, by about 30% for Indian Point 3. This is best seen by comparing Tables 8.5 and 8.6. This comparison shows that the estimates of the parameters have noticeable uncertainty, even if the assumption of a beta distribution is accepted.

Table 8.5 Portion of empirical Bayes analysis, without Kass-Steffey adjustment.

Plant	Beta parameters, ", $	Posterior mean and 90% credible interval [a]
Industry	0.137, 36.34	(6.0E!11, 3.77E!3, 2.12E!2)
Indian Point 3	2.137, 68.34	(6.14E!3, 3.12E!2, 7.15E!2)
Millstone 2	1.137, 47.34	(1.70E!3, 2.40E!2, 6.78E!2)
Crystal River 3	1.137, 52.34	(1.53E!3, 2.17E!2, 6.14E!2)
Indian Point 2	1.137, 60.34	(1.33E!3, 1.88E!2, 5.33E!2)
Robinson 2	1.137, 64.34	(1.24E!3, 1.76E!2, 5.01E!2)
Prairie Isl. 1	0.137, 39.34	(5.5E!12, 3.48E!3, 1.96E!2)
Vogtle 1	0.137, 139.3	(1.6E!12, 9.86E!4, 5.52E!3)

a. Format is (lower bound, mean, upper bound).

Table 8.6 Portion of empirical Bayes analysis, with Kass-Steffey adjustment.

Plant	Beta parameters, ", $	Posterior mean and 90% credible interval [a]
Industry	0.137, 36.34	(6.0E!11, 3.77E!3, 2.12E!2)
Indian Point 3	1.149, 35.65	(2.27E!3, 3.12E!2, 8.77E!2)
Millstone 2	0.596, 24.29	(2.26E!4, 2.40E!2, 8.54E!2)
Crystal R. 3	0.663, 29.94	(3.16E!4, 2.17E!2, 7.44E!2)
Indian Point 2	0.754, 39.34	(4.34E!4, 1.88E!2, 6.17E!2)
Robinson 2	0.793, 44.14	(4.78E!4, 1.76E!2, 5.69E!2)
Prairie Isl. 1	0.133, 37.98	(2.6E!12, 3.48E!3, 1.97E!2)
Vogtle 1	0.127, 128.9	(2.8E!13, 9.86E!4, 5.59E!3)

[a.] Format is (lower bound, mean, upper bound).

Poloski et al. (1998) carry out the above empirical Bayes analysis, but they do not report the calculated lower bounds for plants that experience no failures. They recognize that those lower bounds are unrealistically small, and that such calculated values are an artifact of using a beta distribution. Therefore, they simply report that the lower bound is <1E-8. The next section gives a way to avoid entirely the assumption of a beta distribution.

8.3 The Hierarchical Bayes Method

8.3.1 General Approach

In the preceding discussion of parametric empirical Bayes, the unknown hyperparameters of the population-variability (or prior) distribution were estimated by maximum likelihood. The empirical Bayes estimate of the population-variability distribution is the corresponding distribution in which these maximum likelihood estimates have been inserted.

The hierarchical Bayes method is entirely different. It embodies a complete (or full) Bayesian approach to the problem of estimating the unknown population-variability distribution based on the available data. The hierarchical Bayes approach expresses the initial uncertainty (that is, uncertainty before the data are considered) about the unknown hyperparameters using yet another prior, a so-called second-order or **hyperprior** distribution. For example, in Example 8.1, the population-variability distribution can be a gamma distribution, with parameters (called hyperparameters in this context) α and β. The distribution g could also be lognormal with parameters μ and σ^2. Any desired distribution can be used, with any parameterization. Figure 8.6 denotes the parameters of g generically as α and β. The uncertainty in the state-of-knowledge about the values of α and β is expressed by a suitably specified joint hyperprior distribution on α and β. This expands Figure 8.1 to be Figure 8.6. We almost always desire such hyperprior distributions to be diffuse because we almost never have very precise (or informative) information at the hyperprior level of such a model.

Figure 8.6 is drawn showing α and β with separate distributions. In general, the hyperparameters together have a joint distribution, which does not have to be the product of independent distributions.

In the full Bayesian model all the unknown parameters, including prior-distribution hyperparameters, are assigned prior distributions that express the analyst's initial uncertainty about these parameters. This is known as a **hierarchical Bayes** model. Berger (1985) and Gelman et al. (1995) discuss the basic notions of hierarchical Bayes modeling. In Example 8.1, the parameters of interest to be estimated at the overall population-variability level of the analysis are α and β, while the plant-specific parameters to be estimated are the 66 θ_i values. Each of these 68 parameters is assigned an appropriate prior distribution in a hierarchical Bayes analysis.

The solution to the hierarchical Bayes method requires conditioning on the data and obtaining the required posterior distributions of all the parameters of interest. This is done using Markov chain Monte Carlo (MCMC) simulation (see Section 8.3.3). The desired point and interval estimates of the parameters are then directly (and easily) obtained from these posterior distributions. This process will be illustrated for Examples 8.1 and 8.2 in Sections 8.3.4 and 8.3.5, respectively.

It is well known (Berger 1985 and Gelman et al. 1995) that parametric empirical Bayes can be viewed as an approximation to a full hierarchical Bayes analysis. However, there are several important advantages of hierarchical Bayes over parametric empirical Bayes.

First, parametric empirical Bayes essentially requires the use of a conjugate population-variability distribution in order to obtain the required unconditional distribution of X in closed form. Because hierarchical Bayes analysis is implemented in practice using Monte Carlo simulation, non-conjugate population-variability

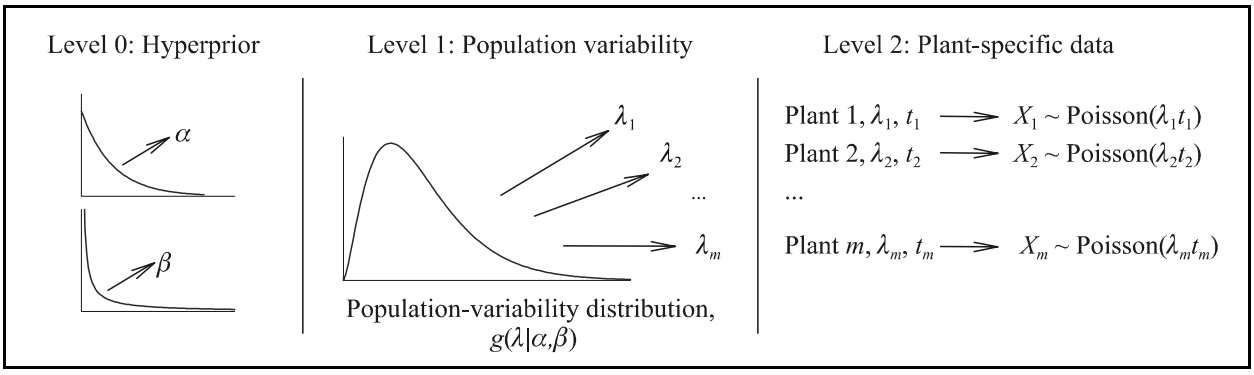


Figure 8.6 Hierarchical Bayes model for Poisson data.

distributions can be used as easily as conjugate distributions. For the Poisson example, a lognormal($\therefore$, F^2) distribution on θ is as easy as a gamma(α, β) distribution.

Second, when using hierarchical Bayes, there is no need to worry about double counting of the data. The hierarchical model and associated Bayesian analysis ensures that this cannot occur.

Finally, as mentioned above, the hierarchical Bayes method is conveniently and easily implemented in practice by means of Markov chain Monte Carlo simulation using existing software, which is presently available for free download from the Web (see Section 8.3.3.3).

8.3.2 Directed Graphs

Those who do not need to know the formulas can skim this section and then jump directly to Section 8.3.4.

The first conceptual step in a hierarchical Bayes analysis should be to construct a **directed graph** representing the hierarchical Bayes model. Briefly, such a graph represents all quantities as nodes in a directed graph, in which arrows between nodes represent directed influences. A directed graph for Example 8.1 is shown in Figure 8.7, where we have defined $\therefore_i = \theta_i t_i$.

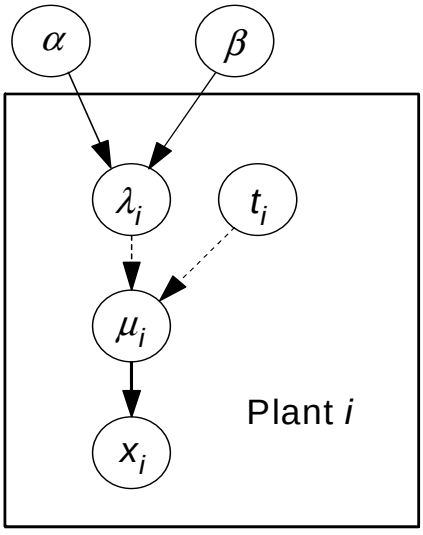


Figure 8.7 Directed graph for the hierarchical Bayes analysis of Example 8.1.

Note that a solid arrow indicates a stochastic dependency, while a dashed arrow indicates a logical function. The hierarchical Bayes approach for the gamma-Poisson case proceeds as follows. First specify hyperprior distributions for the two hyperparameters outside the "plant *i*" box in Figure 8.7. Inference on the hyperparameters α, β and the scram rate vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_{66})$ requires that we obtain Monte Carlo samples from the joint posterior $g(\alpha, \beta, \boldsymbol{\theta} \mid \boldsymbol{x})$, where the data vector $\boldsymbol{x}$ is defined as $\boldsymbol{x} = (x_1, x_2, \ldots, x_{66})$. The letter g is used here to denote both prior and posterior densities. Generate these samples, and then use summary statistics from these samples to obtain the desired estimates, such as point and interval estimates of these parameters.

In order to calculate samples from this joint posterior we must successively sample from the full conditional distributions. That is, we must successively sample from the conditional distribution of each stochastic node given all the other stochastic nodes in the graph. However, conditional independence is expeditiously exploited in directed graphs in order to simplify these full conditional distributions. For example, given β and θ_i, α in Figure 8.7 is conditionally independent of x_i.

8.3.3 Markov Chain Monte Carlo (MCMC) Simulation

Readers who do not need the programming details can skip directly to Section 8.3.4.

Markov chain Monte Carlo (MCMC) sampling techniques give the required samples from the joint posterior distribution of all the unknown parameters. The desired hierarchical Bayes point and interval estimates can thus be directly computed from the corresponding simulated sample observations without the need for tedious analytical or numerical calculations. MCMC is a Monte Carlo integration technique that is implemented using Markov chains. MCMC draws these samples by running a cleverly constructed Markov chain for a long period of time. Good introductions to MCMC are provided by Gilks et al. (1996) and Gelman et al. (1995).

In the Poisson example, the required hierarchical Bayes estimates can be obtained by means of Gibbs sampling, a basic MCMC technique that is described next. The equations are sketched here first. Then a publicly available software package, BUGS, is described. The package implements the equations without requiring the users to understand the details.

8.3.3.1 Gibbs Sampling

Gibbs sampling is a technique that can be used to generate a random sample from the joint posterior distribution indirectly, provided that we can directly sample each of the full conditional distributions (which are described below).

The Gibbs sampling method, sometimes also called simply the **Gibbs sampler**, is briefly described here. Suppose that we have a set of p parameters $\theta_1, \theta_2, \ldots, \theta_p$ whose joint posterior distribution $g(\theta_1, \theta_2, \ldots, \theta_p \mid \mathbf{x})$ is unknown but is of interest to be estimated. This is the usual case when using the hierarchical Bayes method. In Example 8.1, the θ_is consist of the two hyperparameters plus the 66 δ_is, and the number of parameters, p, is 68.

However, suppose that the full conditional distributions $g(\theta_i \mid \theta_j, \mathbf{x}, j \neq i)$ $i = 1, 2, \ldots, p$, are known in the sense that sample values of θ_i, conditional on values of θ_j, $j \neq i$, may be generated from these by some appropriate method. Under mild conditions, these conditional distributions uniquely determine the required joint posterior distribution $g(\theta_1, \theta_2, \ldots, \theta_p \mid \mathbf{x})$; hence, they determine all the unconditional marginal distributions $g(\theta_i \mid \mathbf{x})$, $i = 1, 2, \ldots, p$, as well.

The Gibbs sampler generates samples from the required joint distribution as follows:

(1) Select an arbitrary starting set of values $\theta_1^0, \ldots, \theta_p^0$. Set $j = 0$.

(2) Draw θ_1^{j+1} from $g(\theta_1 \mid \theta_2^j, \ldots, \theta_p^j, \mathbf{x})$, then θ_2^{j+1} from $g(\theta_2 \mid \theta_1^j, \theta_3^j, \ldots, \theta_p^j, \mathbf{x})$, and so on up to θ_p^{j+1} from $g(\theta_p \mid \theta_1^j, \ldots, \theta_{p-1}^j, \mathbf{x})$ to complete one iteration of the sampler.

(3) Increment j and repeat Step (2) until $j+1 = n$. After n such iterations of Step (2), we have obtained the sample $(\theta_1^n, \ldots, \theta_p^n)$. Under mild conditions, as $n \to \infty$ this p-tuple converges in distribution to the unknown joint posterior distribution $g(\theta_1, \theta_2, \ldots, \theta_p \mid \mathbf{x})$.

Typical implementation of the Gibbs sampling algorithm generates an initial "large" number of iterations (called the **burn-in**) until the samples have converged. The burn-in samples are discarded, and the samples generated thereafter are used as sample observations from the joint posterior distribution $g(\theta_1, \theta_2, \ldots, \theta_p \mid \mathbf{x})$. Nonparametric density estimators, such as those given in Section 6.7.3, can then be used to approximate the posterior distribution using the post burn-in samples. Examples 8.1 and 8.2 are analyzed in this way in Sections 8.3.4 and 8.3.5.

In Example 8.1, 68 full conditional distributions are required in order to use the Gibbs sampler:

(1)

$$g(\alpha \mid \beta, \lambda, x) \equiv g(\alpha \mid \beta, \lambda) \propto \left[\frac{\beta^\alpha}{\Gamma(\alpha)}\right]^{66} \left(\prod_{i=1}^{66} \lambda_i\right)^{\alpha} e^{-\alpha}$$

(2)

$$g(\beta \mid \alpha, \delta, x) \;/\; g(\beta \mid \alpha, \delta) = \text{gamma}\left(66\alpha + 0.0625, \sum_{i=1}^{66} \lambda_i + 0.0625\right)$$

(3)

$g(\delta_i \mid \alpha, \beta, x) = \text{gamma}(\alpha + x_i, \beta + t_i)$, $i = 1, \ldots, 66$.

It is easy to sample directly from the gamma distributions. The first distribution, however, the distribution for α, is not of a familiar form, and in fact is not known fully. It is only known to be proportional to the stated expression, with the normalizing constant unknown. The Metropolis-Hastings method is described next, as a way to sample from a distribution such as the distribution for α.

8.3.3.2 Metropolis-Hastings Step

It is sometimes the case that one or more of the full conditional distributions $g(\theta_i \mid \theta_j, \mathbf{x}, j \neq i)$ $i = 1, 2, \ldots, p$, required in Step (2) of the Gibbs sampler may not be available in closed form. This may happen as follows. These full conditional distributions are usually obtained using Bayes' Theorem and, while the two terms (the likelihood and the prior distribution) in the numerator of Bayes' Theorem are usually known, the integration required to obtain the normalizing factor (the denominator) in Bayes' Theorem cannot be performed in closed form. That is, the required full conditional distribution is known only up to a multiplicative constant, the normalizing factor. The corresponding full conditional distribution is thus unavailable in closed form, and sample values from this distribution cannot be directly obtained as required in Step (2) of the Gibbs sampler.

Denote the full conditional distribution, which is known only up to a normalizing constant, as $g(\theta_i \mid \theta_j, \mathbf{x}, j \neq i)$. For convenience, we suppress the conditioning terms in the notation below. In this situation, sample observations may be obtained in Step (2) of the Gibbs sampler

by using a so-called **Metropolis-Hastings step** (Hastings 1970) as follows:

(1) Initialize θ_i^0 and set $j = 0$.
(2) Generate an observation θ_i^* from a candidate distribution $q(\theta_i^* | \theta_i^j)$, where $q(y|x)$ is a probability density in y with parameter (for example mean) x.
(3) Generate a uniform(0, 1) observation u .
(4) Let

$$\theta_i^{j+1} = \begin{cases} \theta_i^*, & \text{if } u \le \alpha(\theta_i^j, \theta_i^*) \\ \theta_i^j, & \text{otherwise} \end{cases}$$

where

$$\alpha(x, y) = \frac{g(y)q(x|y)}{g(x)q(y|x)} .$$

(5) Increment j and go to (2).

Because " uses a *ratio*, $g(y)/g(x)$, it can be calculated even though the normalizing constant for g is unknown. The candidate distribution in Step (2) can be almost any distribution (Gilks et al. 1996), although a symmetric distribution such as a normal distribution results in a simplification of the algorithm, and is called simply a **Metropolis step** (as opposed to a Metropolis-Hastings step). A common choice for $q(y|x)$ is a normal distribution with mean x and a variance that allows the random deviates to be a representative sample from the entire complete conditional distribution. A preliminary rule of thumb suggested by Gelman et al. (1995, Sec. 11.5) is that the variance be such that the new value, θ_i^*, is picked in Step (4) about 30% of the time, and the old value, θ_i^j, is picked about 70% of the time. The new value should be picked more often in problems with few parameters and less often in problems with many parameters.

Actually, BUGS favors a method called **adaptive rejection sampling** (Gilks and Wild 1992) instead of the Metropolis-Hastings algorithm. This method uses more storage space but fewer iterations. It requires that the conditional distributions in the Gibbs sampler be log-concave (George et al. 1993). This requirement is satisfied for the commonly used prior distributions. If the user happens to select a prior that leads to a problem, BUGS will give a diagnostic message.

8.3.3.3 BUGS (Bayesian Inference Using Gibbs Sampling)

Fortunately, for a wide range of common problems, there is little need to actually program the Gibbs sampler in practice. Gibbs sampling has been conveniently implemented through the **BUGS** software project (BUGS 1995, Spiegelhalter et al. 1995, and Gilks et al. 1994). It is currently available for free download at WWW URL

http://www.mrc-bsu.cam.ac.uk/bugs/ .

The classic BUGS program uses text-based model description and a command-line interface, and versions are available for major computing platforms.

A Windows version, WinBUGS, has an optional graphical user interface (called DoodleBUGS) as well as on-line monitoring and convergence diagnostics. BUGS is reasonably easy to use and, along with a user manual, includes two volumes of examples. Section 8.3.4 illustrates how WinBUGS was used in Example 8.1 to obtain the hierarchical Bayes estimates of the 66 plant-specific scram rates and of the corresponding population-variability distribution. Section 8.3.5 uses WinBUGS in Example 8.2, to obtain the hierarchical Bayes estimates of p at each of the 68 plants and of the corresponding population-variability distribution.

8.3.4 Application to Example 8.1, Initiating Events

8.3.4.1 Development of Hyperprior Distributions

It will be necessary at the start to develop hyperprior distributions. Therefore, this topic is discussed here in general terms. One must choose hyperpriors that are diffuse enough to contain the true values of the hyperparameters. That means that the analyst must have some vague idea of the true values of the hyperparameters. In principle, this should come from experience with other, similar data sets. Often, however, the current data set is used to provide guidance on the plausible hyperparameter values. The diffuseness of the hyperpriors can be assessed by using the rough rule of thumb that a quantity can plausibly vary from its mean by up to two standard deviations.

In the examples below, the plausible values will be obtained from the previous empirical Bayes analyses. However, it is possible to develop hyperprior distributions without such a sophisticated analysis, just by examining the data carefully. For example, one can easily get an estimate of the mean of β, and one can also find an upper bound that β rarely exceeds. Such considerations can provide rough guidance on plausible bounds for the hyperparameters " and $. Then one constructs hyperprior distributions that achieve these

bounds without going far into the tails of the distributions. The process is not always easy, but some analysis of this type is necessary.

It might seem tempting to just use distributions with enormous standard deviations, say mean 1 and standard deviation 1,000. In principle this is possible, but in practice it may challenge the numerical accuracy of the software. Therefore, it is recommended that the data be examined and that hyperpriors be constructed that are diffuse enough to include anything consistent with the data, but that are not unrealistically diffuse.

8.3.4.2 Example Analysis

Let us now illustrate the hierarchical Bayes method in Example 8.1. We use the Gibbs sampler in BUGS to calculate all the required population and plant-specific scram rate estimates.

To begin, we assume that the population-variability distribution is gamma(″, $) just as in Section 8.2. First the hyperprior distributions must be constructed for ″ and $, based on the considerations in Section 8.3.4.1. The empirical Bayes analysis of Section 8.2.5.1 gave ″ = 1.4 and $ = 1.2. Therefore, we choose both hyperpriors to have mean 1, agreeing (to one significant digit) with the empirical Bayes results. The hyperparameter ″ is given an exponential hyperprior distribution with mean and variance of 1, while the hyperparameter $ is given an independent gamma(0.0625, 0.0625) hyperprior distribution. Thus, $ is assumed to have a hyperprior mean and standard deviation of 1 and 4, respectively. The hyperpriors are diffuse (large variances), and have plausible means, so they will probably not bias the final answers much.

Figure 8.8 contains the WinBUGS model used here for this Poisson example. The initial values considered are: alpha = 1, beta = 1, and lambda[i] = 1, i = 1, …, 66.

After 1,000 burn-in iterations (to remove the effect of the initial starting values and to achieve convergence of the Markov chain), 10,000 additional simulated posterior sample values of ″, $, and 8 = ($_1$, $\_2$, …, $\_{66}$) were recorded. These 10,000 sample values were then used to calculate the required posterior point and credible interval estimates of ″, $, and each $\_i$. For example, the hierarchical Bayes estimated posterior mean of the Callaway scram rate is calculated to be 4.97 per 1,000 critical hours. The corresponding 90% credible interval on $\_1$ is [2.87, 7.51].

In addition, the marginal posterior mean and standard deviation of ″ are calculated to be 1.38 and 0.30, respectively, whereas those for $ are computed to be 1.21 and 0.32. A hierarchical Bayes 90% credible interval for ″ is [0.94, 1.93], while the corresponding interval for $ is [0.76, 1.78]. The marginal posterior correlation between ″ and $ is also easily calculated from the 10,000 pairs of corresponding posterior (″, $) values to be 0.89. From Table 8.3 we see that the empirical Bayes point estimates of ″ and $ are 1.39 and 1.21, respectively, which are in near perfect agreement with the hierarchical Bayes estimates.

```
model
{
  for (i in 1:M) {
    lambda[i] ~ dgamma(alpha,beta)
    mu[i] <- lambda[i]*t[i]
    x[i] ~ dpois(mu[i])
  }
  alpha ~ dexp(1.0)
  beta ~ dgamma(0.0625, 0.0625)
}
```

Figure 8.8 WinBUGS model specification for Example 8.1.

Figure 8.9 illustrates the hierarchical Bayes results. For each plant, the posterior mean and 90% credible interval are shown. The mean and 90% credible interval for the population-variability distribution are also shown, and a vertical dashed line is drawn through the mean. Actually, this population-variability distribution is the gamma distribution evaluated when ″ and $ are set equal to their posterior means. It does not reflect the uncertainty in these two hyperparameters. Figure 8.9 agrees very closely with Figure 8.3.

Table 8.7 contains the same portion of the numerical hierarchical Bayes analysis results for Example 8.1 as are displayed in Table 8.3.

The point and interval estimates in Table 8.7 are all in good agreement with the empirical Bayes estimates in Table 8.3.

8.3.5 Application to Example 8.2, AFW Segment Failures to Start

Recall that this example has 68 plants with sparse failure data consisting of only 6 failures in 1,993 demands. Because the data are so sparse, the form of the prior, the

population-variability distribution, can strongly influence the answers. Therefore, the example is analyzed using two population-variability distributions, first a beta distribution, as in Section 8.2, and then a logistic-normal distribution. In each case, diffuse hyperpriors with plausible means are used. Therefore, the exact choices made for the hyperpriors have little influence on the answer.

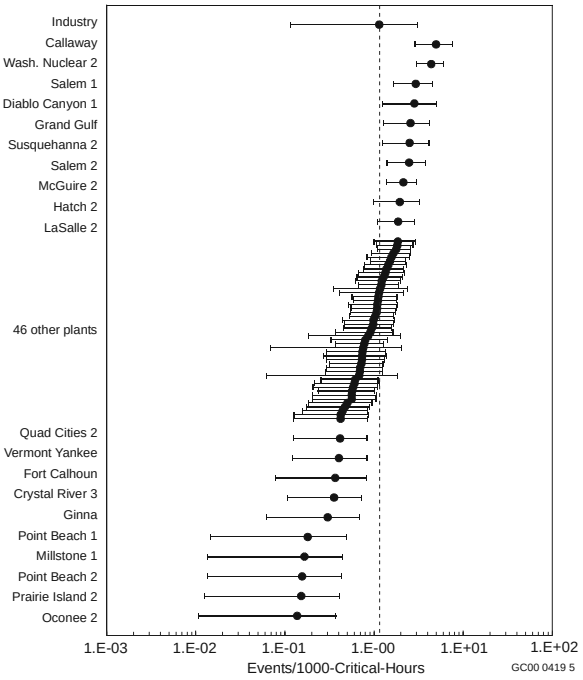


Figure 8.9 Plant-specific posterior means and 90% credible intervals for δ, from hierarchical Bayes analysis.

Table 8.7 Portion of hierarchical Bayes analysis results for Example 8.1.

Plant	Posterior mean and 90% credible interval [a]
Industry	(0.116, 1.14, 3.06)
Callaway	(2.87, 4.97, 7.51)
Wash. Nuc. 2	(3.01, 4.39, 5.99)
Salem 1	(1.65, 2.94, 4.57)
Diablo Can. 1	(1.24, 2.83, 4.99)
Grand Gulf	(1.27, 2.57, 4.21)
Pt. Beach 2	(0.013, 0.156, 0.429)
Prairie Isl. 2	(0.012, 0.152, 0.410)
Oconee 2	(0.011, 0.137, 0.374)

[a.] Format is (lower bound, mean, upper bound). The least significant digit may be inaccurate by 2 or more, because of Monte Carlo sampling error.

8.3.5.1 Analysis with Beta Prior

We assume that the population-variability distribution is a beta(α, β) distribution. In this case, the empirical Bayes analysis found α = 0.1 and β = 36. We will construct diffuse hyperpriors that contain these values.

The hyper-parameter α is assigned an exponential (1) hyperprior distribution with a hyperprior mean and standard deviation of 1, while the hyperparameter β is assigned an independent gamma(1.0, 0.035) hyperprior distribution. Thus, we assume that β has a hyperprior mean and standard deviation both equal to approximately 30. The forms of these hyperprior distributions ensure that the joint posterior distribution will be log-concave, and the diffuseness of the hyperpriors ensures that they will not influence the final answers greatly.

The chosen hyperpriors include the desired values of 0.1 and 36, and much more. In particular, the exponential distribution for α allows any value below the mean of 1, because the exponential density decreases monotonically. Likewise, the gamma density of β also is monotonically decreasing, and so allows any value below the mean. As for values larger than the mean, we apply the rule of thumb that says that most random variables can easily deviate from the mean by up to two standard deviations. This says that α could be as large as 3 and β as large as about 90. Because of the knowledge gained from the empirical Bayes analysis, we believe that the hyperpriors are diffuse enough. If the prior belief about the parameters were more rough, based on a cruder analysis of the data, we would want to compensate by choosing more diffuse hyperpriors. See Section 8.3.4.1 for more details.

Fig. 8.10 contains the WinBUGS model used here for this binomial example. The initial values were: alpha = 1, beta = 1, and p[i] = 0.01, i = 1, …, 68.

```
model
{
  for (i in 1:M) {
     x[i] ~ dbin(p[i],n[i])
     p[i] ~ dbeta(alpha, beta)
  }
  alpha ~ dexp(1.0)
  beta  ~ dgamma(1.0,0.035)
}
```

Figure 8.10 WinBUGS model specification for a beta prior in Example 8.2.

After 10,000 burn-in iterations (to remove the effect of the initial starting values and to achieve convergence of the Markov chain), 90,000 additional simulated posterior sample values of ″, $, and **p** = ($p_1$, p_2, …, p_{68}) were recorded. These 90,000 sample values were then used to calculate the required posterior point and credible interval estimates of ″, $, and each p_i.

In addition, the marginal posterior mean and standard deviation of ″ are calculated to be 0.188 and 0.136, respectively, whereas those for $ are computed to be 46.4 and 32.4. A hierarchical Bayes 90% credible interval for ″ is [0.046, 0.442], while the corresponding interval for $ is [9.95, 109.5]. Note the large uncertainties associated with the hierarchical Bayes estimates of ″ and $ because of the sparseness of the data. Table 8.6 shows that the empirical Bayes point estimates of ″ and $ are 0.137 and 36.34, respectively, which are well within the 90% credible intervals of the corresponding hierarchical Bayes estimates.

Figure 8.11 illustrates the hierarchical Bayes results. For each plant, the posterior mean and 90% credible interval are shown. The mean and 90% credible interval for the population-variability distribution are also shown, and a vertical dashed line is drawn through the mean.

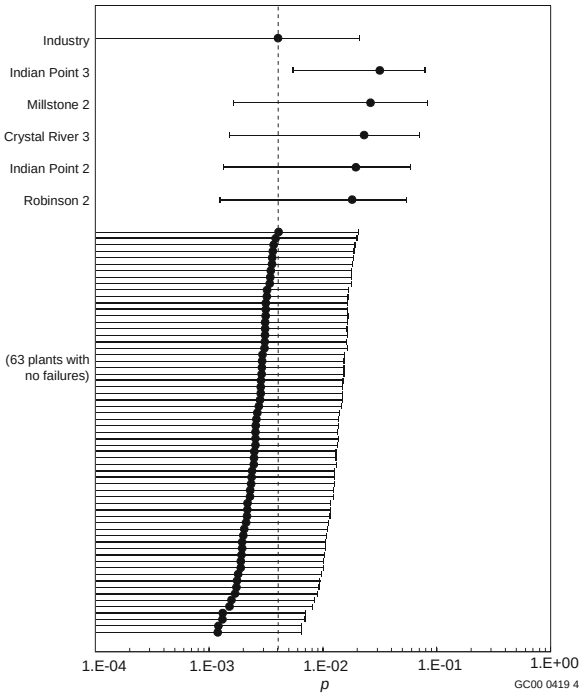


Figure 8.11 Plant-specific posterior means and 90% credible intervals for *p*, from hierarchical Bayes analysis with beta population-variability distribution.

Table 8.8 contains the same portion of the numerical hierarchical Bayes analysis results for Example 8.2 as given in Table 8.6. The results are presented to only two significant digits, because the Monte Carlo errors reported by BUGS show that the third significant digit is not meaningful.

Table 8.8 Portion of hierarchical Bayes analysis results using beta prior for Example 8.2.

Plant	Posterior mean and 90% credible interval [a]
Industry	(1.6E!9, 4.0E!3, 2.1E!2)
Indian Point 3	(5.5E!2, 3.2E!2, 7.9E!2)
Millstone 2	(1.6E!2, 2.6E!2, 8.3E!2)
Crystal R. 3	(1.5E!2, 2.3E!2, 7.1E!2)
Indian Point 2	(1.3E!2, 1.9E!2, 5.9E!2)
Robinson 2	(1.2E!2, 1.8E!2, 5.4E!2)
Prairie Isl. 1	(1.5E!15, 4.1E!3, 2.1E!2)
Vogtle 1	(2.5E!16, 1.2E!3, 6.4E!3)

[a] Format is (lower bound, mean, upper bound).

The point and interval estimates in Table 8.8 are all in reasonably close agreement with the empirical Bayes estimates in Table 8.6.

8.3.5.2 Analysis with Logistic-Normal Prior

One of the primary advantages in using the hierarchical Bayes method is the ability to consider non-conjugate population-variability (or prior) distributions. We now illustrate this for Example 8.2.

The previous analysis considered a conjugate beta prior in this example. Table 8.8 shows that, for $x = 0$, the use of a beta prior produces lower 5% credible limits on the order of 10^{-15} or 10^{-16}, which are unrealistically small. This result is a consequence of the fitted L-shaped beta prior distribution, with high density close to p = 0.

We can avoid such unrealistic results by using a non-conjugate logistic-normal prior distribution (see Section 6.3.2.5.2 or A.7.9) in the hierarchical Bayes approach. Recall that, while such a prior is extremely difficult to consider in an empirical Bayes approach, it is extremely easy to do in hierarchical Bayes.

Figure 8.12 contains the WinBUGS model specification for using a logistic-normal prior in Example 8.2. Observe that this is no more difficult than using the

hierarchical Bayes model based on a conjugate beta prior shown in Figure 8.10. Note that BUGS parameterizes the normal distribution in terms of $J = 1/F^2$. A commonly used prior distribution for J is gamma, and that choice is used here. Thus, assigning $:$ a prior precision of 0.0001 is equivalent to assigning it a prior variance of 10,000, or a prior standard deviation of 100.

```
model
{
   for (i in 1:M) {
      x[i] ~ dbin(p[i],n[i])
      y[i] ~ dnorm(mu,tau)
      p[i] <- exp(y[i])/(1 + exp(y[i]))
   }
   mu ~ dnorm(-5,0.0001)
   tau ~ dgamma(1,7)
}
```

Figure 8.12 WinBUGS model specification for a logistic-normal prior in Example 8.2.

Again using 10,000 burn-in iterations and 90,000 replications of the Gibbs sampler for the model in Figure 8.12, WinBUGS likewise calculated posterior means and 90% credibility intervals for $:$, F, and each p_i.

The marginal posterior mean and standard deviation of $:$ are calculated to be –5.097 and 0.09517, respectively, whereas those for F are computed to be 0.640 and 0.238. A hierarchical Bayes 90% credible interval for $:$ is [-5.253, -4.939], while the corresponding interval for F is [0.322, 1.08].

Figure 8.13 shows the two estimated population-variability distributions, when the form is assumed to be beta (the conjugate distribution) or logistic-normal. The mean of the beta prior is 0.004 and the mean of the logistic-normal prior is 0.007, nearly twice as large. Note that, unlike the beta prior, the logistic-normal prior avoids the high probability density close to $p = 0$.

Figure 8.14 illustrates the hierarchical Bayes results using the logistic-normal prior. As in Figure 8.11, the posterior mean and 90% credible interval are shown for each plant. The mean and 90% credible interval for the population-variability distribution are also shown, and a vertical dashed line is drawn through the mean. This plot may be directly compared with Figure 8.11.

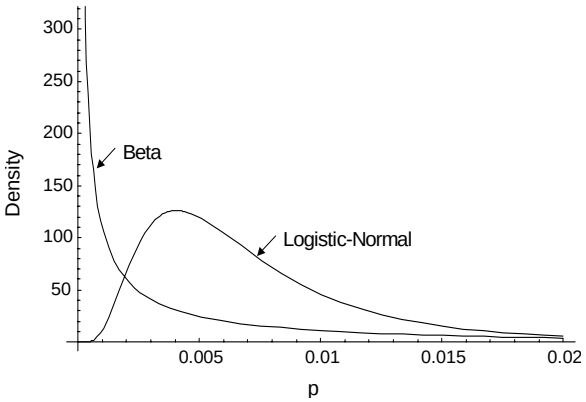


Figure 8.13 Fitted population-variability distributions in Example 8.2.

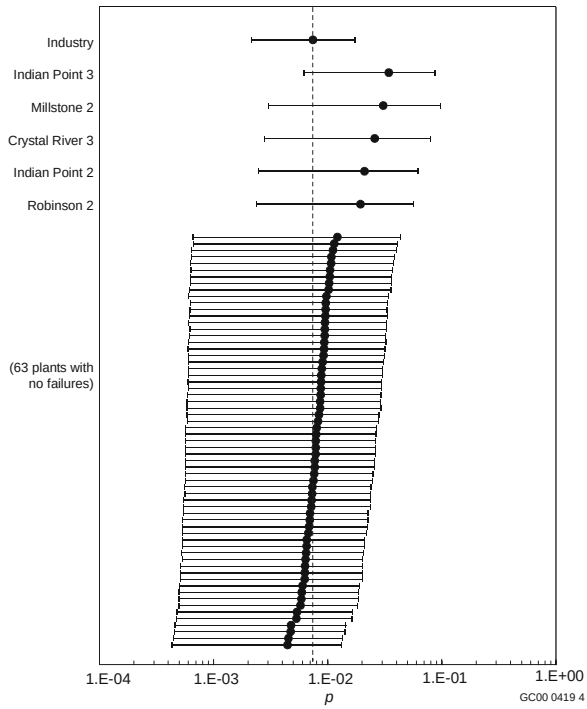


Figure 8.14 Plant-specific posterior means and 90% credible intervals for p, from hierarchical Bayes analysis with logistic-normal population-variability distribution.

Table 8.9 contains the same portion of the numerical hierarchical Bayes analysis results for the logistic-normal as given in Table 8.8 for the beta prior.

Note that the posterior means and 90% credible intervals in Table 8.9 are all larger than those in Table 8.8. As stated above, this is a direct consequence of using the logistic-normal prior in Figure 8.13. Observe also that the lower 90% credibility limits for plants with no failures, such as Prairie Island 1 and Vogtle 1, are now much more realistic than the corresponding limits in Table 8.8.

Table 8.9 Portion of hierarchical Bayes analysis results using logistic-normal prior for Example 8.2.

Plant	Posterior mean and 90% credible interval [a]
Industry	(2.1E!3, 7.4E!3, 1.7E!2)
Indian Point 3	(6.2E!3, 3.4E!2, 8.7E!2)
Millstone 2	(3.0E!3, 3.1E!2, 9.7E!2)
Crystal R. 3	(2.8E!3, 2.6E!2, 8.0E!2)
Indian Point 2	(2.5E!3, 2.1E!2, 6.2E!2)
Robinson 2	(2.4E!3, 1.9E!2, 5.7E!2)
Prairie Isl. 1	(6.6E!4, 1.2E!2, 4.4E!2)
Vogtle 1	(4.3E!4, 4.E!3, 1.3E!2)

[a.] Format is (lower bound, mean, upper bound).

8.4 Discussion

This chapter concludes with several important observations.

8.4.1 Hierarchical Bayes Is Still Bayes

The hierarchical Bayes model is a special case of the familiar Bayesian model. It is not some new kind of construction. To see this, consider Levels 0 and 1 together in Figure 8.6. The prior parameter is a vector $\mathbf{2}$, consisting of the hyperparameters and the θ_is. Thus, the prior $\mathbf{2}$ is a vector with dimension 2+ *m*. The prior distribution on $\mathbf{2}$ is specified: the joint distribution of the hyperparameters α and β is given by the hyperprior in Level 0, and the conditional distributions of the θ_is are independent, and specified by *g*, conditional on α and β. Thus, Levels 0 and 1 together specify the prior parameter vector and its prior distribution. The posterior distribution is therefore given by Bayes' Theorem:

$$g_{post}(\mathbf{2} \mid \text{data}) \propto \Pr(\text{data} \mid \mathbf{2}) \times g_{prior}(\mathbf{2}) .$$

This differs from the applications of Bayes' Theorem elsewhere in this handbook in only two ways: the parameter is a high-dimensional vector, and the prior distribution has a lot of structure, as shown in Figure 8.6.

A practical consequence of the high dimension of $\mathbf{2}$ is that the tools of Chapter 6, numerical integration and simple random sampling methods, do not work well. More recently developed methods, versions of Markov chain Monte Carlo sampling, must be used. Conceptually, however, hierarchical Bayes modeling fits perfectly within the framework of Bayes' Theorem. In particular, everything is legal, with no double counting of data.

8.4.2 The "Two-Stage" Bayesian Method

Kaplan (1983) introduced a "two-stage" Bayesian method, which has sometimes been used in PRA work. It is described here in terms of Figure 8.6. The method singles out the plant of special interest, say Plant 1. It then estimates the hyperparameters, α and β, in a Bayesian way, using the data from *all the plants except Plant 1*. It then uses the estimated $g(\theta \mid \alpha, \beta)$ as a prior, combining it with the data from Plant 1 to estimate θ_1 in the usual Bayesian way.

The originally intended reason for not using Plant 1 in the first stage was to avoid double counting. As mentioned in Section 8.4.1, the hierarchical Bayes method is based directly on Bayes' Theorem, and therefore does not involve double counting. Therefore, the two-stage Bayesian method should no longer be used, but should be replaced by the conceptually cleaner hierarchical Bayes method. Now that numerical algorithms have been developed to sample from the posterior distributions, this is completely feasible.

8.4.3 Lower Bounds on Parameters

Example 8.2 illustrated that different population-variability distributions (prior distributions) can lead to radically different lower percentiles of the parameters' posterior distributions. This occurred in that example for those plants that had experienced no failures. A beta prior led to 5th percentiles for *p* on the order of 1E!15, whereas a logistic-normal prior led to 5th percentiles on the order of 5E!4. No one believes the first answers, but many people could believe the second answers. Does that mean that the answers from the logistic-normal prior are "right"?

In fact, the lower bounds are an artifact of the model, in both cases. The plants that experienced no failures reveal only that *p* is "small" at those plants. They do not give information about how small *p* might be. If many plants have no failures, as in Example 8.2, then we have very little information about the lower end of the population-variability distribution. In contrast to this, the plants that experienced one or more failures convey much more information, revealing both how large and how small *p* could plausibly be at those

plants. Therefore, the 95th percentile of p at any plant is *somewhat* dependent on the assumed form of the population-variability distribution (beta, logistic-normal, or whatever). But when many plants have no observed failures, the 5th percentile of p at any of those plants is *extremely* dependent on this assumed form.

And why was a particular form assumed for the population-variability distribution? For convenience only! Thus, even if the answers from a logistic-normal prior look credible, we do not "know" that they are correct. We may choose to discard the results from using a beta prior, because we do not want to publish 5th percentiles that could be ridiculed. We might also choose to publish the results from using a logistic-normal prior, knowing that the 5th percentiles appear credible. But it is a delusion to think that we "know" lower bounds on p at the plants with no observed failures. The calculated lower bounds remain an artifact of the assumed model.

Fortunately, lower bounds are not a concern for risk. Means and upper bounds are the important quantities, and they can be estimated with much less dependence on the model.

8.4.4 Empirical Bayes as an Approximation to Hierarchical Bayes

As remarked elsewhere, Figures 8.3 and 8.9 are very similar to each other, and Figures 8.5 and 8.11 are similar to each other. That is, in both Examples 8.1 and 8.2, the empirical Bayes results are numerically close to the hierarchical Bayes results, when (a) the empirical Bayes method includes the Kass-Steffey adjustment, and (b) both methods use the conjugate population-variability distribution, a gamma distribution for Poisson data and a beta distribution for binomial data. This agreement between the methods is more than coincidence. Kass and Steffey (1989) developed their method specifically with this intent: to make the empirical Bayes approach give a first-order approximation to the hierarchical Bayes approach with a diffuse hyperprior. The method does this well in the two examples. Of course, when the hierarchical Bayes method does not use a conjugate population-variability distribution, as in Section 8.3.5.2, there is no corresponding empirical Bayes method.

9. REFERENCES

ANS and IEEE, 1983, *PRA Procedures Guide A Guide to the Performance of Probabilistic Risk Assessments for Nuclear Power Plants*, NUREG/CR-2300, 1983, American Nuclear Society, LaGrange, IL, and The Institute of Electrical and Electronics Engineers, New York, NY.

ASME, 2002, *Standard for Probabilistic Risk Assessment for Nuclear Power Plant Applications*, ASME RA-S-2002. American Society of Mechanical Engineers, New York, NY.

Ascher, H., and H. Feingold, 1984, *Repairable Systems Reliability - Modeling, Inference, Misconceptions and Their Causes*. Marcel Dekker, Inc., NY.

ATV, 1992, *T-Book*, 3rd ed., ATV Office and Studsvik, Vattenfall AB, S-16287 Vällingby, Sweden.

Atwood, C. L., 1992, Parametric Estimation of Time-Dependent Failure Rates for Probabilistic Risk Assessment in *Reliability Engineering and System Safety*, Vol. 37, No. 3, pp. 181-194.

Atwood, C. L., 1994, *Hits per Trial: Basic Analysis of Binomial Data*, EGG-RAAM-11041, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Atwood, C. L., 1995, *Modeling Patterns in Count Data Using Loglinear and Related Models*, INEL-95/012, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Atwood, C. L., 1996, Constrained Noninformative Priors in Risk Assessment, in *Reliability Engineering and System Safety*, Vol. 53, Issue 1, pp. 37-46.

Atwood, C. L., and C. D. Gentillon, 1996, Bayesian Treatment of Uncertainty in Classifying Data: Two Case Studies, in *Probabilistic Safety Assessment and Management 96 (ESREL'96 -- PSAM-III)*, P. C. Cacciabue and I. A. Papazoglou, ed., pp. 1283-1288. Springer, London.

Atwood, C.L., D. L. Kelly, F. M. Marshall, D. A. Prawdzik, and J. W. Stetkar, 1998, *Evaluation of Loss of Offsite Power Events at Nuclear Power Plants: 1980-1996*, NUREG/CR-5496, INEEL/EXT-97-00887, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Atwood, C. L., and M. Engelhardt, 2003, *Bayesian Estimation of Unavailability*, STATWOOD 2003-2. Submitted to *Reliability Engineering and System Safety*.

Bain, L. J., and M. Engelhardt, 1991, *Statistical Analysis of Reliability and Life-Testing Models, Theory and Methods*, 2nd ed. Marcel Dekker, Inc., NY.

Bain, L. J., and M. Engelhardt, 1992, *Introduction to Probability and Mathematical Statistics*, 2nd ed. PWS-Kent Pub., Boston, MA.

Battle, R. E., and D. J. Campbell, 1983, *Reliability of Emergency AC Power Systems at Nuclear Power Plants*, NUREG/CR-2989, Oak Ridge National Laboratory, Oak Ridge, TN.

Berger, J. O., 1985, *Statistical Decision Theory and Bayesian Analysis*. 2nd ed., Springer-Verlag, NY.

Bernardo, J. M., and A. F. M. Smith, 1994, *Bayesian Theory*, John Wiley and Sons, New York.

Bertucio, R. C., and J. A. Julius, 1990, *Analysis of Core Damage Frequency: Surry, Unit 1 Internal Events*, NUREG/CR-4550,Vol. 3, Rev. 1, Part 1, SAND86-2084, Sandia National Laboratories, Albuquerque, NM.

Blanton, C. H., and S. A. Eide, 1993, *Savannah River Site Generic Data Base Development*, WSRC-TR-93-262, Westinghouse Savannah River Company, Aiken, SC.

Box, G. E. P., 1980, Sampling and Bayes' Inference in Scientific Modelling and Robustness (with discussion) in *Journal of the Royal Statistical Society, Series A*, Vol. 143, Part 4, pp. 383-430.

Box, G. E. P., and G. C. Tiao, 1973, *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading, MA.

Breeding, Roger J., Timothy J. Leahy, and Jonathan Young, 1985, *Probabilistic Risk Assessment Course Documentation Volume 1: PRA Fundamentals*, NUREG/CR-4350 Volume 1 of 7, SAND85-1495 Volume 1 of 7, Sandia National Laboratories, Albuquerque NM, August 1985.

Budnitz, R.J., et al. 1997. *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on the Use of Experts*, NUREG/CR-6372, Lawrence Livermore National Laboratory, Livermore, CA.

BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.50, 1995, MRC Biostatistics Unit, Cambridge, UK. Available at www.mrc-bsu.cam.ac.uk/bugs/.

Carlin, B. P., and T. A. Louis, 2000, *Bayes and Empirical Bayes Methods for Data Analysis, Second Edition*, Chapman & Hall/CRC, Boca Raton, FL.

Carlson, D. D., D. R. Gallop, A. M. Kolaczkowski, G. J. Kolb, and D. W. Stack, 1983, *Interim Reliability Evaluation Program Procedures Guide*, NUREG/CR-2728, SAND82-1100, Sandia National Laboratories, Albuquerque, NM.

Chu, T.L., et al., 1993, *Evaluation of Potential Severe Accidents During Low Power and Shutdown Operations at Surry Unit 1*, NUREG/CR-6144, BNL-NUREG-52399, Brookhaven National Laboratory, Upton, NY.

Cleveland, W. S., 1985, *The Elements of Graphing Data*. Wadsworth Advanced Books and Software, Monterey, CA.

Com Ed, 1981, *Zion Probabilistic Safety Study*, Commonwealth Edison, Chicago, IL.

Conover, W. J. 1999. *Practical Nonparametric Statistics, Third Edition*. John Wiley & Sons, NY.

Cox, D. R., and D. V. Hinkley, 1974, *Theoretical Statistics*, Chapman and Hall, London, England.

Cox, D. R., and P. A. W. Lewis, 1966, *The Statistical Analysis of Series of Events.* Chapman and Hall, London, England.

CPC, 1981, *Big Rock Point Nuclear Power Plant Probabilistic Risk Assessment*, Consumers Power Company, Jackson, MI.

Denson, G., G. Chandler, W. Crowell, A. Clark, and P. Jaworski, 1995, *Nonelectronic Parts Reliability Data - 1995*, NPRD-95. Reliability Analysis Center, Griffiss AFB, Rome, NY.

Denson, W., W. H. Crowell, D. Mahar, and P. Jaworski, 1996, *Electronics Parts Reliability Data - 1997, A Compendum of Commerical and Military Device Field Failure Rates: Volumes 1 and 2*, EPRD-97, Reliability Analysis Center, Rome, NY.

Drago, J. P., R. J. Borkowski, D. H. Pike, and F. F. Goldberg, 1982, *The In-Plant Reliability Data Base for Nuclear-Power-Plant Components: Data Collection and Methodology Report*, NUREG/CR-2641, Oak Ridge National Laboratory, Oak Ridge, TN.

Droguett, E., 1999, *A Methodology for the Treatment of Model Uncertainty*, PhD Dissertation, Dept. of Materials and Nuclear Engineering, University of Maryland, College Park, MD.

Drouin, M. T., F. T. Harper, and A. L. Camp, 1990, *Analysis of Core Damage Frequency From Internal Events, Methodology Guidelines: Volume 1*, NUREG/CR-4550, SAND-86-2084. Sandia National Laboratories, Albuquerque, NM.

Duke, 1984, *Oconee Unit III Probabilistic Risk Assessment*. Duke Power Company, Charlotte, NC.

Eide, S. A., S. T. Beck, M. B. Calley, W. J. Galyean, C. D. Gentillon, S.T. Khericha, S.D. Novack, and T.E. Wierman, 1999a, *Reliability Study: Westinghouse Reactor Protector System, 1984-1995*, NUREG/CR-5500,Vol. 2, INEEL/EXT-97-00740, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Eide, S. A., S. T. Beck, M. B. Calley, W. J. Galyean, C. D. Gentillon, S.T. Khericha, and T.E. Wierman, 1999b, *Reliability Study: General Electric Reactor Protector System, 1984-1995*, NUREG/CR-5500,Vol. 3, INEEL/EXT-97-00740, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Eide, S. A., S. V. Chmielewski, and T. D. Swantz, 1990, *Generic Component Failure Data Base for Light Water and Liquid Sodium Reactor PRAs (Probabilistic Risk Assessments)*, EGG-SSRE-8875, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Eide, S. A., S. T. Khericha, M. B. Calley, D. A. Johnson, and M. L. Marteeny, 1991, *Component External Leakage and Rupture Frequency Estimates*, EGG-SSRE-9639, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Engelhardt, M. E., 1994, *Events in Time: Basic Analysis of Poisson Data*, EGG-RAAM-11088, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Engelhardt, M. E., 1995, Reliability Estimation and Applications, chap. 6 in *The Exponential Distribution: Theory, Methods, and Applications*, N. Balakrishnan and A. P. Basu, ed., pp. 77-91. Gordon and Breach Publishers, Amsterdam, Holland.

Engelhardt, M. E., 1996, *Statistical Analysis of Random Duration Times*, INEL-95/0206, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

EPRI, 1982, *ATWS: A Reappraisal, Part 3: Frequency of Anticipated Transients (Interim Report)*, EPRI NP-2230. Electric Power Research Institute, Palo Alto, CA.

EPRI, 1989, *Advanced Light Water Reactor Requirements Document*, Electric Power Research Institute, Palo Alto, CA.

EPRI, 1992, *Nuclear Plant Reliability: Data Collection and Usage Guide*, EPRI TR-100381, Halliburton NUS Environmental Corporation, Walnut Creek, CA.

Everitt, B. S., 1992, *The Analysis of Contingency Tables*, 2nd ed. Chapman and Hall, London, England.

Federal Register Notice for revisions to 10 CFR 50.72 and 10 CFR 50.73, October 2000.

Fleming, K.N., and A. Mosleh, 1985, *Classification and Analysis of Reactor Operating Experience Involving Dependent Events*, EPRI NP-3967, Electric Power Research Institute, Palo Alto, CA.

Gallaher, R.B., et al., 1984, *Sequence Coding and Search System for Licensee Event Reports*, NUREG/CR-3905, ORNL/NSIC-223, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 1995. *Bayesian Data Analysis*. Chapman & Hall, London.

George, E. I., U. E. Makov, and A. F. M. Smith, 1993, Conjugate Likelihood Distributions, in *Scandinavian Journal of Statistics*, Vol. 20, pp. 147-156.

Gertman, D. I., W. E. Gilmore, W. J. Galyean, M. R. Groh, and C. D. Gentillon, 1990, *Nuclear Computerized Library for Assessing Reactor Reliability (NUCLARR). Volume 1, Revision 1. Summary Description*, NUREG/CR-4639, EFF-2458-VOL-1-REV-1, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter, 1996, *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

Gilks, W. R., A. Thomas, and D. J. Spiegelhalter, 1994, A Language and Program for Complex Bayesian Modeling, in *The Statistician*, Vol. 43, pp. 169-178.

Gilks, W. R., and P. Wild, 1992, Adaptive Rejection Sampling for Gibbs Sampling, in *Applied Statistics*, Vol. 41, pp. 337-348.

Grant, G.M., W. S. Roesener, D. G. Hall, C. L. Atwood, C. D. Gentillon, and T. R. Wolf, 1999a, *Reliability Study: High-Pressure Coolant Injection (HPCI) System, 1987-1993*, NUREG/CR-5500, Vol. 4, INEL-94/0158, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Grant, G.M., J. P. Poloski, A. J. Luptak, C. D. Gentillon, and W. J. Galyean, 1999b, *Reliability Study: Emergency Diesel Generator Power System 1987-1993*, NUREG/CR-5500, Vol. 5, INEL-95/0035, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Grant, G.M., J. P. Poloski, C. D. Gentillon, and W. J. Galyean, 1999c, *Reliability Study: Isolation Condenser System 1987-1993*, NUREG/CR-5500, Vol. 6, INEL-95/0478, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Hagler-Bailly, 1972, *Nuclear Power Experience Database*. Hagler-Bailly Inc., Boulder, CO.

Hart, Jeffrey D., 1997, *Nonparametric Smoothing and Lack-of-Fit Tests*, Springer, New York.

Hastings, W. K., 1970, Monte Carlo Sampling Methods Using Markov Chains and Their Applications, in *Biometrika*, Vol. 57, pp. 97-109.

Hatch, S. W., P. Cybulskis, and R. O. Wooton, 1981, *Reactor Safety Study Methodology Applications Program: Grand Gulf #1 BWR Power Plant*, NUREG/CR-1659/4, Sandia National Laboratories, Albuquerque, NM.

Hoeting, Jennifer A., David Madigan, Adrian E. Raftery, and Chris T. Volinsky, 1999, Bayesian Model Averaging: A Tutorial, in *Statistical Science*, Vol. 14, No. 4, pp. 382-417.

Hollander, M., and D. A. Wolfe, 1999, *Nonparametric Statistical Methods*, 2nd ed. John Wiley & Sons, NY.

Houghton J. R., 2001a, *Component performance study – air-operated valves, 1987 – 1998*, NUREG-1715, Vol. 3, Nuclear Regulatory Commission, Washington, DC.

Houghton J. R., 2001b, *Component performance study – motor-operated valves, 1987 – 1998*, NUREG-1715, Vol. 4, Nuclear Regulatory Commission, Washington, DC.

Houghton J. R. and H. G. Hamzehee, 2000a, *Component performance study – turbine-driven pumps, 1987 – 1998*, NUREG-1715, Vol. 1, Nuclear Regulatory Commission, Washington, DC.

Houghton J. R., and H. G. Hamzehee, 2000b, *Component performance study – motor-driven pumps, 1987 – 1998,* NUREG-1715, Vol. 2, Nuclear Regulatory Commission, Washington, DC.

Høyland, A., and M. Rausand. 1994. *System Reliability Theory*. Wiley-Interscience, John Wiley & Sons, NY.

Hubble, W. H., and C. Miller, 1980, *Data Summaries of Licensee Event Reports of Valves at U. S. Commercial Nuclear Power Plants*, NUREG/CR-1363, Vol. 1, EGG-EA-5125, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

IEEE, 1983, *IEEE Guide to the Collection and Presentation of Electrical, Electronic, Sensing Component, and Mechanical Equipment Reliability Data for Nuclear Power Generating Stations*, IEEE-Std 500-1984. Institute of Electrical and Electronics Engineers, NY.

INPO, 1998, *Equipment Performance and Information Exchange System (EPIX), Volume1, Instructions for Data Entry Maintenance Rule and Reliability Information Module*, Vol. 1, INPO 98-001, Institute of Nuclear Power Operations, Atlanta, GA.

Johnson, N. L., S. Kotz, and A. W. Kemp, 1992, *Univariate Discrete Distributions*, 2nd ed. John Wiley & Sons, NY.

Johnson, N. L., S. Kotz, and N. Balakrishnan, 1994, *Continuous Univariate Distributions, Volume 1*, 2nd ed. John Wiley & Sons, NY.

Johnson, N. L., S. Kotz and N. Balakrishnan, 1995, *Continuous Univariate Distributions, Volume 2*, 2nd ed. John Wiley & Sons, NY.

Kaplan, S., 1983, On a "Two-Stage" Bayesian Procedure for Determining Failure Rates from Experiential Data, in *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-102, No. 1, pp. 195-202.

Kass, R. E., and D. Steffey. 1989. Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models), in *Journal of the American Statistical Association*, Vol. 84, pp 717-726.

Koehler, K. J., and K. Larntz, 1980, An Empirical Investigation of Goodness-of-Fit Statistics for Sparse Multinomials, in *Journal of the American Statistical Association*, Vol. 75, Issue 370, pp. 336-344.

Kolaczkowski, A. M., and A. C. Payne, 1983, *Station Blackout Accident Analysis*, NUREG/CR-3226, SAND-82-2450, Sandia National Laboratories, Albuquerque, NM.

Kotra, J.P., M.P. Lee, N.A. Eisenberg, and A.R. DeWispelare. 1996. *Branch Technical Position on the Use of Expert Elicitation in the High-Level Radioactive Waste Program*, NUREG/CR-1563, U.S. Nuclear Regulatory Commission, Washington, DC.

Kvarfordt, K.J., M.J. Cebull, S.T. Wood, and A. Mosleh, 1998, *Common-Cause Failure Database and Analysis System: Overview*, NUREG/CR-6268, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Lee, Peter M., 1997, *Bayesian Statistics: An Introduction, Second Edition*, Arnold, a member of the Hodder Headline Group, London.

Lehtinen, E., U. Pulkkinen, and K. Pörn, 1997, *Statistical Trend Analysis Methods for Temporal Phenomena*, SKI Report 97:10, Swedish Nuclear Power Inspectorate, Stockholm, Sweden.

Lewontin, R. C., and J. Felsenstein, 1965, The Robustness of Homogeneity Tests in $2 \times N$ Tables, in *Biometrics*, No. 21, pp. 19-33.

Lofgren, E. V., and M. Thaggard, 1992, *Analysis of Standby and Demand Stress Failures Modes: Methodology and Applications to EDGs and MOVs*, NUREG/CR-5823, SAIC 87-029, Science Applications International Corporation.

Lwin, T., and N. Singh, 1974, Bayesian Analysis of the Gamma Distribution Model in Reliability Engineering, in *IEEE Transactions on Reliability*, Vol. R-23, pp. 314-319.

Mackowiak, D. P., C. D. Gentillon, and K. L. Smith, 1985, *Development of Transient Initiating Event Frequencies for Use in Probabilistic Risk Assessments*, NUREG/CR-3862, EGG-2323, EG&G Idaho, Inc., Idaho Falls, ID.

Manning, J. J., et al., 1986, *A Guide for Completing Input Data Forms for CREDO, A Centralized Reliability, Availability and Maintainability Analysis Center for Liquid-Metal-Cooled Reactor and Test Facility Components*, ORNL/TM-9892, Oak Ridge National Laboratory, Oak Ridge, TN.

Marshall, F.M., D.M. Rasmuson, and A. Mosleh, 1998, *Common-Cause Parameter Estimations*, NUREG/CR-5497, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Martz, H.F., and R.G. Almond, 1997, Using Higher-Level Failure Data in Fault Tree Quantification, in *Reliability Engineering and System Safety*, Vol. 56, pp. 29-42.

Martz, H. F., and M. C. Bryson, 1983, On Combining Data for Estimating the Frequency of Low-Probability Events with Application to Sodium Valve Failure Rates, in *Nuclear Science and Engineering*, Vol. 83, No. 2, pp. 267-280.

Martz, H. F., and M. S. Hamada, 2003, Uncertainty in Counts and Operating Time in Estimating Poisson Occurrence Rates, in *Reliability Engineering and System Safety*, Vol. 80, pp. 75-79.

Martz, H. F., R. L. Parker, and D. M. Rasmuson, 1999, Estimation of Trends in the Scram Rate at Nuclear Power Plants, in *Technometrics*, Vol. 41, pp. 352-364.

Martz, H. F., and R. R. Picard, 1995. Uncertainty in Poisson Event Counts and Exposure Time in Rate Estimation, in *Reliability Engineering and System Safety*, Vol. 48, pp. 181-190.

Martz, H. F., P. H. Kvam, and L. R. Abramson, 1996. Empirical Bayes Estimation of the Reliability of Nuclear-Power-Plant Emergency Diesel Generators, in *Technometrics*, Vol. 38, pp. 11-24.

Martz, H. F., and R. A. Waller, 1991, *Bayesian Reliability Analysis*. R. E. Krieger Publishing Co., Malabar, FL.

McClymont, A. S., and B. W. Poehlman, 1982a, *ATWS: A Reappraisal, Part 3: Frequency of Anticipated Transients, Interim Report*, EPRI-NP-2230-PT.3, Electric Power Research Institute, Palo Alto, CA.

McClymont, A. S., and B. W. Poehlman, 1982b, *Loss of Off-site Power at Nuclear Power Plants: Data and Analysis*, EPRI NP-2301, Electric Power Research Institute, Palo Alto, CA.

McCullagh, P., and J. A. Nelder, 1989, *Generalized Linear Models*, 2nd ed. Chapman & Hall, NY.

Microsoft Excel, 2001, Microsoft Corporation, Seattle, WA.

Moore, D. S., 1986, Tests of Chi-Squared Type, chap. 3 in *Goodness-of-Fit Techniques*, Ralph B. D'Agostino and Michael A. Stephens, ed., pp. 63-95. Marcel Dekker, Inc., NY.

Mosleh, A., 1986, Hidden Sources of Uncertainty: Judgment in the Collection and Analysis of Data, in *Nuclear Engineering and Design*, Vol. 93, pp. 187-198.

Mosleh, A., and G. Apostolakis, 1985, The Development of a Generic Data Base for Failure Rates, in Proceedings of the *ANS/ENS International Topical Meeting on Probabilistic Safety Methods and Applications*, held in San Francisco, CA on February 24 - March 1, EPRI NP-3912-SR.

Mosleh, A., K. N. Fleming, G. W. Parry, H. M. Paula, D. H. Worledge, and D. M. Rasmuson, 1988, *Procedures for Treating Common Cause Failures in Safety and Reliability Studies*, NUREG/CR-4780, EPRI NP-5613, PLG-0547, Vol. 1.

Mosleh, A., K. N. Fleming, G. W. Parry, H. M. Paula, and D. H. Worledge, 1989, *Procedures for Treating Common Cause Failures in Safety and Reliability Studies: Analytical Background and Techniques*, NUREG/CR-4780V2, PLG-0547-VOL-2, EPRINP-5613-VOL-2, Pickard, Lowe, and Garrick, Inc., Newport Beach, CA.

Mosleh, A., D.M. Rasmuson, and F.M. Marshall, 1998, *Guidelines on Modeling Common-Cause Failures in Probabilistic Risk Assessments*, NUREG/CR-5485, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Mosleh, A., N. Siu, C. Smidts, and C. Lui, 1994, *Proceedings of Workshop I in Advanced Topics in Risk and Reliability Analysis. Model Uncertainty: Its Characterization and Quantification*, NUREG/CP-1038.

Nelson, W., 1982, *Applied Life Data Analysis*. John Wiley & Sons, NY.

Neter, J., and W. Wasserman, 1974, *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*. Richard D. Irwin, Inc., Homewood, IL.

Nickolaus, J.R., N. E. Moffitt, B. F. Gore, and T. V. Vo, 1992, *Auxiliary Feedwater System Risk-Based Inspection Guide for the North Anna Nuclear Power Plants*, NUREG/CR-5837, PNL-7923, Battelle Pacific Northwest Laboratory, Richland, WA.

Nowlen, S. P., A. Boone, A. Kolaczkowski, D. Kelly, N. Siu, and H. W. R. Woods, 2002, Fire Detection and Suppression: PRA Modeling and Data Analysis, in *Proceedings of the 6th International Conference on Probabilistic Safety Assessment and Management*. Elseveir, 2002, Volume I, pgs 243-249.

NRC, 1975a, *Reporting of Operating Information - Appendix A Technical Specifications*, Draft Regulatory Guide 1.16, Nuclear Regulatory Commission, Washington, D.C.

NRC, 1975b, *Reactor Safety Study - An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants*, WASH-1400, NUREG-75/014, Nuclear Regulatory Commission, Washington, D.C.

NRC, 1990, *Severe Accident Risks: An Assessment for Five U.S. Nuclear Power Plants: Final Summary Report*, NUREG-1150-V1, Nuclear Regulatory Commission, Washington, D.C.

NRC, 1995a, Use of Probabilistic Risk Assessment Methods in Nuclear Activities: Final Policy Statement, *Federal Register*, Vol. 60, pp. 42622. Washington, D. C.

NRC, 1995b, *Licensed Operating Reactors – Status Summary Report*, NURG-0020, Nuclear Regulatory Commission, Washington, D.C.

NRC, 1997, *Revised Contents of the Monthly Operating Report*, Generic Letter 97-02, Nuclear Regulatory Commission, Washington, D.C.

NRC, 1998, *Event Reporting Guidelines 10 CFR 50.72 and 50.73*, NUREG-1022, Rev. 1, Nuclear Regulatory Commission, Washington, D.C.

NRC, 2000a, *Event Reporting Guidelines 10 CFR 50.72 and 50.73*, NUREG-1022, Rev. 2, Nuclear Regulatory Commission, Washington, D.C.

NRC, 2000b, *Technical Basis and Implementation Guidelines for A Technique for Human Event Analysis*, NUREG-1624, Rev. 1, Nuclear Regulatory Commission, Washington, D.C.

OREDA, 1997, *Offshore Reliability Data Handbook*, OREDA-97, 3rd edition, SINTEF Industrial Management, Det Norske Veritas, Veritasveien 1, P.O. Box 300, –1322 Høvik, Norway.

Oswald, A. J., C. D. Gentillon, S. D. Matthews, and T. R. Meachum, 1982, *Generic Data Base for Data and Models Chapter of the National Reliability Evaluation Program (NREP) Guide*, EGG-EA-5887, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Papazoglou, I. A., R. A. Bari, A. J. Buslik, R. E. Hall, and D. Ilberg, 1984, *Probabilistic Safety Analysis Procedures Guide*, NUREG/CR-2815, Brookhaven National Laboratory, Upton, NY.

Parry, G.W., B.O.Y. Lydell, A.J. Spurgin, P. Moieni, and A. Beare, 1992, *An Approach to the Analysis of Operator Actions in Probabilistic Risk Assessment*, EPRI TR-100259, Electric Power Research Institute, Palo Alto, CA.

PECO, 1982, *Probabilistic Risk Assessment: Limerick Generation Station, Revision 4*, Philadelphia Electric Company, Philadelphia, PA.

Poloski, J. P., and W. H. Sullivan, 1980, *Data Summaries of Licensee Event Reports of Diesel Generators at U.S. Commercial Nuclear Power Plants, January 1, 1976 to December 31, 1978*, NUREG/CR-1362, EGG-EA-5092, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Poloski, J. P., G. M. Grant, C. D. Gentillon, W. J. Galyean, and J. K. Knudsen, 1998, *Reliability Study: Auxiliary/Emergency Feedwater System, 1987-1995*, NUREG/CR-5500, Vol. 1, INEEL/EXT-97/00740, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Poloski, J.P., D. G. Marksberry, C. L. Atwood, and W. J. Galyean, 1999a, *Rates of Initiating Events at U.S. Nuclear Power Plants: 1987-1995*, NUREG/CR-5750, INEEL/EXT-98-00401, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Poloski, J. P., G. M. Grant, C. D. Gentillon, W. J. Galyean, and W. S. Roesener, 1999b, *Reliability Study: Reactor Core Isolation Cooling System, 1987-1993*, NUREG/CR-5500, Vol. 7, INEL-95/0196, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Poloski, J. P., G. M. Grant, C. D. Gentillon, and W. J. Galyean, 1999c, *Reliability Study: High Pressure Core Spray System, 1987-1993*, NUREG/CR-5500, Vol. 8, INEL-95/00133, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Poloski, J. P., J. K. Knudsen, C.L. Atwood, and W. J. Galyean, 2000, *Reliability Study: High Pressure Safety Injection System, 1987-1997*, NUREG/CR-5500, Vol. 9, INEEL/EXT-99-00373. Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Pörn, K., 1990, *On Empirical Bayesian Inference Applied to Poisson Probability Models*, Linköping Studies in Science and Technology, Dissertation No. 234. Division of Quality Technology, Department of Mechanical Engineering, Linköping University, Linköping, Sweden, ISBN 91-7890-696-3.

Pörn, K., 1996, "The two-stage Bayesian Method used for the T-Book Application," in *Reliability Engineering and System Safety*, Vol. 51, Issue 2, pp. 169-179.

Pörn, K., R. Blomquist, and K. Shen, 1993, Initiating Events in Swedish Nuclear Power Plants – Classification and Trends, in *Proceedings of the International Topical Meeting on Probabilistic Safety Assessment, PSA '93*, clearwater Beach, FL. Published by the American Nuclear Society, ISBN 0-89448-180-0, pp. 1190-1194.

Pörn, K., K. Shen, and R. Nyman, 1994, *I-Book, Version 2. Initiating Events at the Nordic Nuclear Power Plants* (in Swedish), SKI Report 94:12, Swedish Nuclear Power Inspectorate, Stockholm, Sweden. ISSN 1104-1374, ISRN SKI-R-94-12-SE.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992, *Numerical Recipes in Fortran 77: Volume 1 of Fortran Numerical Recipes, Second Edition*, Cambridge University Press, New York. (Also available for C++ and Pascal.)

Quattro Pro, 2001, Corel Corporation, Ottawa, Canada.

Raiffa, H., and R Schlaifer, 1961, *Applied Statistical Decision Theory*, The M.I.T. Press, Cambridge, MA.

Roscoe, J. T., and J. A. Byars, 1971, An Investigation of the Restraints with Respect to Sample Size Commonly Imposed on the Use of the Chi-square Statistic, in *Journal of the American Statistical Association*, Vol. 66, Issue 336, pp. 755-759.

Ross, S., 1983, *Stochastic Processes*, John Wiley and Sons, NY.

Rubin, D. B., 1996, Multiple Imputation after 18+ Years (with Discussion), in *Journal of the American Statistical Association*, Vol. 91, pp. 444-489.

Samanta, P. K., I. S. Kim, T. Mankamo, and W. E. Vesely, 1994, *Handbook of Methods for Risk-Based Analyses of Technical Specifications*, NUREG/CR-6141, BNL-NUREG-52398, Brookhaven National Laboratory, Upton, NY.

Sandtorv, H. A., P. Hokstad, and D. W. Thompson, 1996, Practical Experiences with a Data Collection Project: The OREDA project, in *Reliability Engineering and System Safety*, Vol. 51, Issue 2, pp. 159-167.

SAS, Version 8. 2000. SAS Institute, Cary, NC.

SAS/INSIGHT Users Guide, Version 6, Third Edition, 1995, SAS Institute Inc., SAS Campus Drive, Cary, NC.

Shah, V. N., M. B. Sattison, C. L. Atwood, A. G. Ware, and G. M. Grant, 1998, *Assessment of Pressurized Water Reactor Primary System Leaks*, NUREG/CR-6582, INEEL/EXT-97-01068, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Silverman, B. W., 1986, *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, England.

Simard, R. L., 1983, Recent Initiatives in Information Exchange Through the International Nuclear Community, in *Progress in Nuclear Energy*, Vol. 12, No. 3, pp. 221-225.

Simard, R. L., 1985, Application of NPRDS Data to Enhance Nuclear Plant Design and Operations, in *Proceedings: International Topical Meeting on Probabilistic Safety Methods and Applications Volume: Sessions 1-8*, conference held in Palo Alto, CA, on Feb. 24 - March 1, EPRI NP-3912-SR, Volume 1, Electric Power Research Institute, Palo Alto, CA.

Siu, N., and G. Apostolakis, 1984, *Models for the Detection and Suppression of Fires in Nuclear Power Plants and Some Related Statistical Problems*, UCLA-ENG-8440.

Siu, N., and G. Apostolakis, 1986, A Methodology for Analyzing the Detection and Suppression of Fires in Nuclear Power Plants, in *Nuclear Science and Engineering*, Vol. 94, pp. 213-226.

Siu, N. O., and D. L. Kelly, 1998, Bayesian Parameter Estimation in Probabilistic Risk Assessment, in *Reliability Engineering and System Safety*, Volume 62, Issue 1-2, pp. 89-116.

Smith, A. F. M., and A. E. Gelfand, 1992, Bayesian Statistics without Tears: A Sampling-Resampling Perspective, in *The American Statistician*, Vol. 46, No. 2, pp. 84-88.

Spiegelhalter, D. J., A. Thomas, N. G. Best, and W. R. Gilks, 1995, *BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.50*, MRC Biostatistics Unit, Cambridge, UK. Available at www.mrc-bsu.cam.ac.uk/bugs/

Staple, B.D., T.D. Brown, J.J. Gregory, and J.L. LaChance, 1999, *Risk Comparisons of Scheduling Preventative Maintenance for Boiling Water Reactors During Shutdown and Power Operation*, NUREG/Cr-5593, SAND99-0676, Sandia National Laboratories, Albuquerque, NM.

STATISTICA for Windows, Vols. I and III, 1995, StatSoft, Inc., 23255 East 13th Street, Tulsa, OK.

StatXact, 1999, CYTEL Software Corp., Cambridge, MA.

Swain, A.D, 1987, *Accident Sequence Evaluation Program Human Reliability Analysis Procedure*, NUREG/CR-4772, Sandia National Laboratories, Albuquerque, NM.

Swain, A.D., and H.E. Guttman, 1983, *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications*, NUREG/CR-1278, Sandia National Laboratories, Albuquerque, NM.

Tashjian, B. M., 1982, Nuclear Plant Reliability Data System Data Usage in Availability Engineering, in *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-101, No. 8, pp. 2582-2590.

Thompson, W. A., Jr., 1981, On the Foundations of Reliability, in *Technometrics*, Vol. 23, No. 1, pp. 1-13.

Trojovsky, M., 1982, *Data Summaries of Licensee Event Reports Of Pumps at U. S. Commercial Nuclear Power Plant*, NUREG/CR-1205, Revision 1, EGG-EA-5524, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

TUD Office and Pörn Consulting, 2000, *T-Book: Reliability Data of Components in Nordic Nuclear Power Plants, 5th Edition*, Battenfall Inhouse-Grafiska, Vällingby, Sweden.

Tukey, J. W., 1977, *Exploratory Data Analysis*. Addison-Wesley Publishing Co., Reading, MA.

Vesely, W.E., F. F. Goldberg, N. H. Roberts, and D. F. Haasl, 1981, *Fault Tree Handbook*, NUREG-0492, U. S. Nuclear Regulatory Commission, January 1981.

Walpole, R. E., and R. H. Myers, 1985, *Probability and Statistics for Engineers and Scientists*, 2nd edition. Macmillan, NY.

Wheeler, T.A., S. C. Hora, W. R. Crammand, and S. D. Unwin, 1989, *Analysis of Core Damage Frequency from Internal Events: Expert Judgment Elicitation*, NUREG/CR-4550, Vol. 2, SAND86-2084, Sandia National Laboratories, Albuquerque, NM.

Whitehead, D.W., H.M. Paula, G.W. Perry, and D.M. Rasmuson, 1993, *Enhancements to Data Collection and Reporting of Single and Multiple Failure Events*, NUREG/CR-5471, SAND89-2562, Sandia National Laboratories, Albuquerque, NM.

Wierman, T.E., S.T. Beck, M.B. Calley, S.A. Eide, C.D. Gentillon, and W.E. Kohn, 2002a, *Reliability Study: Combustion Engineering Reactor Protection System, 1984-1998*, NUREG/CR-5500, Vol. 10, INEL/EXT-97-00740, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Wierman, T.E., S.T. Beck, M.B. Calley, S.A. Eide, C.D. Gentillon, and W.E. Kohn, 2002b, *Reliability Study: Babcock & Wilcox Reactor Protection System, 1984-1998*, NUREG/CR-5500, Vol. 11, INEL/EXT-97-00740, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Wilk, M. B., and R. Gnanadesikan, 1968, Probability Plotting Methods for the Analysis of Data, in *Biometrika*, Vol. 55, pp. 1-17.

Zacks, S., 1992, *Introduction to Reliability Analysis: Probability Models and Statistical Methods*, Springer-Verlag, New York.

APPENDICES

A. BASICS OF PROBABILITY

A.1 Events

Any repeatable process for which the result is uncertain can be considered an experiment, such as counting failures over time or measuring time to failure of a specific item of interest. The result of one execution of the experiment is referred to as an **outcome**. Due to uncertainty associated with the process, repetitions or trials of a defined experiment would not be expected to produce the same outcomes. The set of all possible outcomes of an experiment is defined as the **sample space**.

Sample spaces can contain discrete points (such as pass, fail) or points in a continuum (such as measurement of time to failure). An **event** E is a specified set of possible outcomes in a sample space S (denoted $E \subseteq S$, where $\subseteq$ denotes subset).

Most events of interest in practical situations are compound events, formed by some composition of two or more events. Composition of events can occur through the union, intersection, or complement of events, or through some combination of these.

For two events, E_1 and E_2, in a sample space S, the **union** of E_1 and E_2 is defined to be the event containing all sample points in E_1 or E_2 or both, and is denoted by the symbol $(E_1 \cup E_2)$. Thus, a union is simply the event that either E_1 or E_2 occurs or both E_1 and E_2 occur.

For two events, E_1 and E_2, in a sample space S, the **intersection** of E_1 and E_2 is defined to be the event containing all sample points that are in both E_1 and E_2, denoted by the symbol $(E_1 \cap E_2)$. The intersection is the event that both E_1 and E_2 occur.

Figure A.1 shows a symbolic picture, called a **Venn diagram**, of some outcomes and events. In this example, the event E_1 contains three outcomes, event E_2 contains five outcomes, the union contains seven outcomes, and the intersection contains one outcome.

The **complement** of an event E is the collection of all sample points in S and not in E. The complement of E is denoted by the symbol $\bar{E}$ and is the outcomes in S that are not in E occur. In Figure A.1, the complement of E_1 is an event containing seven outcomes.

It is sometimes useful to speak of the **empty** or **null** set, a set containing no outcomes. In Figure A.1, the event E_3 is empty. It cannot occur.

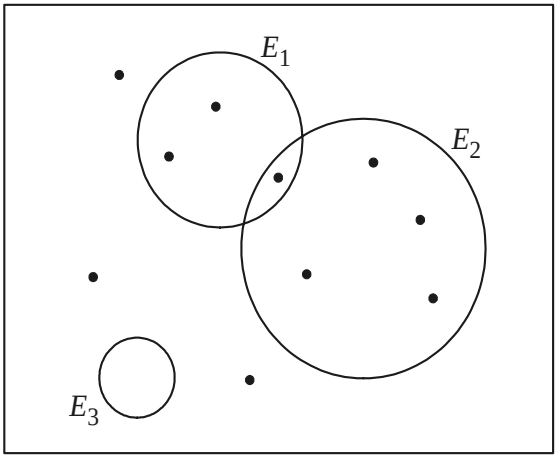


Figure A.1 Venn diagram, showing ten outcomes and three events.

Two events, E_1 and E_2, are said to be **mutually exclusive** if the event $(E_1 \cap E_2)$ contains no outcomes in the sample space S. That is, the intersection of the two events is the null set. Mutually exclusive events are also referred to as **disjoint** events. Three or more events are called mutually exclusive, or disjoint, if each pair of events is mutually exclusive. In other words, no two events can happen together.

A.2 Basic Probability Concepts

Each of the outcomes in a sample space has a probability associated with it. **Probabilities of outcomes** are seldom known; they are usually estimated from relative frequencies with which the outcomes occur when the experiment is repeated many times. Once determined, the probabilities must satisfy two requirements:

1. The probability of each outcome must be a number ≥ 0 and ≤ 1.

2. The probabilities of all outcomes in a given sample space must sum to 1.

Associated with any event E of a sample space S is the **probability of the event**, Pr(E). Since an event represents a particular set of outcomes of an experiment, the values of Pr(E) are built from the probabilities of the outcomes in E.

Probabilities are associated with each outcome in the sample space through a **probability model**. Probability

models are often developed on the basis of information derived from outcomes obtained from an experiment. Probability models are also formulated in the context of mathematical functions.

The values of Pr(E) estimated from the experimental outcomes are often defined as being representative of the *long-run relative frequency* for event E. That is, the relative frequency of an outcome will tend toward some number between 0 and 1 (inclusive) as the number of repetitions of the experiment increases. Thus, the probability of the outcome is the number about which the long-term relative frequency tends to stabilize.

This interpretation forms the basis of the **relative frequency definition of probability**, also referred to as the **frequentist view** of probability. In the frequentist view, a mathematical theory of probability is developed by deriving theorems based on the axioms of probability given in the next subsection. The probability of an event is considered to be a fixed quantity, either known or unknown, that is a property of the physical object involved and that can be estimated from data. A theorem derived from the three axioms describes the frequentist view:

If an experiment is repeated a large number of times, n, the observed relative frequency of occurrence, n_E/n, of the event E (where n_E = the number of repetitions when event E occurred) will tend to stabilize at a constant, Pr(E), referred to as the probability of E.

Another interpretation of probability leads to the so-called **classical definition of probability**, which can be stated as follows:

If an experiment can result in n equally likely and mutually exclusive outcomes and if n_E of these outcomes contain attribute E, then the probability of E is the ratio n_E / n.

For example, if each of the outcomes in Figure A.1 had equal probability, 0.1, then Pr(E_1) = 0.3, Pr(E_2) = 0.5, Pr($E_1 \cap E_2$) = 0.1, Pr($E_1 \cup E_2$) = 0.7, and Pr(E_3) = 0.

The classical definition is limited, because it assumes equally likely outcomes. However, it helps motivate the frequentist axioms mentioned above. These axioms provide a mathematical framework for probability, an overview of which is addressed in Section A.3. Some texts, including parts of this handbook, use the terms *classical* and *frequentist* interchangeably.

Another interpretation of probability is as a **subjective probability**. Probabilities obtained from the opinions of people are examples of subjective probabilities. In this concept, probability can be thought of as a rational measure of belief. Any past information about the problem being considered can be used to help assign the various probabilities. In particular, information about the relative frequency of occurrence of an event could influence the assignment of probabilities.

The notion of subjective probability is the basis for Bayesian inference. In contrast to the relative frequency definition of probability that is based on properties of events, subjective probability can be extended to situations that cannot be repeated under identical conditions. However, the assignment of subjective probabilities can be done according to certain principles so that the frequency definition requirements of probability are satisfied. All the mathematical axioms and theorems developed for frequentist probability apply to subjective probability, but their interpretation is different.

Martz and Waller (1991) present subjective probability as dealing not only with events but with propositions. A proposition is considered to be a collection of events that cannot be conceived as a series of repetitions, for example, a nuclear power plant meltdown. The degree of belief in proposition A, Pr(A), represents how strongly A is believed to be true. Thus, subjective probability refers to the degree of belief in a proposition. At the extremes, if A is believed to be true, Pr(A) = 1; if A is believed to be false, Pr(A) = 0. Points between 0 and 1 represent intermediate beliefs between false and true.

A.3 Basic Rules and Principles of Probability

The relative frequency, classical, and subjective probability definitions of probability satisfy the following axiomatic requirements of probability:

If Pr(E) is defined for a type of subset of the sample space S, and if

1. Pr(E) $\geq$ 0, for every event E,
2. Pr($E_1 \cup E_2 \cup \ldots$) = Pr(E_1) + Pr(E_2) + $\ldots$, where the events E_1, E_2, . . . , are such that no two have a point in common, and
3. Pr(S) = 1,

then Pr(E) is called a **probability function**.

A probability function specifies how the probability is distributed over various subsets E of a sample space S.

From this definition, several rules of probability follow that provide additional properties of a probability function.

The probability of an impossible event (the empty or null set) is zero, written as:

$$\Pr(\emptyset) = 0,$$

where $\emptyset$ is the null set. The probability of the complement of E is given by:

$$\Pr(\bar{E}) = 1 - \Pr(E).$$

In general, the probability of the union of any two events is given by:

$$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2).$$

If E_1 and E_2 are mutually exclusive, then $\Pr(E_1 \cap E_2) = \Pr(\emptyset) = 0$, and

$$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2),$$

which is a special case of the second axiom of probability stated above and is sometimes referred to as the addition rule for probabilities.

For three events,

$$\begin{aligned}\Pr(E_1 \cup E_2 \cup E_3) = {} & \Pr(E_1) + \Pr(E_2) + \Pr(E_3) \\ & - \Pr(E_1 \cap E_2) - \Pr(E_1 \cap E_3) \\ & - \Pr(E_2 \cap E_3) + \Pr(E_1 \cap E_2 \cap E_3).\end{aligned}$$

This rule is also referred to as the **inclusion-exclusion principle** and can be generalized to n events. It is widely used in PRA to calculate the probability of an "or" gate (a union of events) in a fault tree (NRC 1994).

The inclusion-exclusion principle also provides useful upper and lower bounds on the probability of the union of n events that are *not* mutually exclusive. One such upper bound, referred to as the **rare event approximation**, is:

$$\Pr(E_1 \cup E_2 \cup \ldots \cup E_n) \le \Pr(E_1) + \Pr(E_2) + \ldots + \Pr(E_n).$$

The rare event approximation should only be used when the probabilities of the n events are all very small (NRC 1994). If the n events are mutually exclusive, the error is zero. A bound on the error is

$$\binom{n}{2} \max [\Pr(E_i)],$$

which is valid regardless of the independence of events (NRC 1994, though printed with a misprint there). The error in the rare-event approximation arises from the remaining terms in the full expansion of the left-hand side of the inequality. This approximation is frequently used in accident sequence quantification.

Many experimental situations arise in which outcomes are classified by two or more events occurring simultaneously. The simultaneous occurrence of two or more events (the intersection of events) is called a joint event, and its probability is called a **joint probability**. Thus, the joint probability of both events E_1 and E_2 occurring simultaneously is denoted by $\Pr(E_1 \cap E_2)$.

The probability associated with one event, irrespective of the outcomes for the other events, can be obtained by summing all the joint probabilities associated with all the outcomes for the other events, and is referred to as a **marginal probability**. A marginal probability is therefore the *unconditional* probability of an event, unconditioned on the occurrence of any other event.

Two events E_1 and E_2 are often related in such a way that the probability of occurrence of one depends on whether the other has or has not occurred. The **conditional probability** of one event, given that the other has occurred, is equal to the joint probability of the two events divided by the marginal probability of the given event. Thus, the conditional probability of event E_2, given event E_1 has occurred, denoted $\Pr(E_2 * E_1)$, is defined as:

$$\Pr(E_2 * E_1) = \Pr(E_1 \cap E_2) / \Pr(E_1), \qquad \text{(A.1)}$$

for $\Pr(E_1) > 0$. If $\Pr(E_1) = 0$, $\Pr(E_2 * E_1)$ is undefined.

Rearranging this equation yields:

$$\begin{aligned}\Pr(E_1 \cap E_2) &= \Pr(E_1) \Pr(E_2 * E_1) \\ &= \Pr(E_2) \Pr(E_1 * E_2).\end{aligned} \qquad \text{(A.2)}$$

Calculation of joint probability requires the concept of **statistical independence**. An event E_2 is statistically independent of E_1 if the probability of E_2 does not change whenever E_1 occurs or does not occur. Thus, E_2 is **independent** of E_1 if

$$\Pr(E_2 * E_1) = \Pr(E_2).$$

It follows from Equation A.1 that E_2 is independent of E_1 if their joint probability is equal to the product of the *unconditional*, or *marginal*, probabilities of the events:

$$\Pr(E_1 \cap E_2) = \Pr(E_1) \Pr(E_2).$$

This is sometimes referred to as the multiplication rule for probabilities. In this formulation, it is clear that E_2 is independent of E_1 if E_1 is independent of E_2, and we say simply that E_1 and E_2 are statistically independent. If $\Pr(E_2)$ varies depending on whether or not event E_1 has occurred, then events E_1 and E_2 are said to be **statistically dependent**.

If E_1, E_2, ... are mutually exclusive, and if the union of E_1, E_2, ... equals the entire sample space, then the events E_1, E_2, ... are said to form a **partition** of the sample space. Exactly one of the events must occur, not more than one but exactly one. In this case, the **law of total probability** says

$$\Pr(A) = \sum \Pr(A \mid E_i) \Pr(E_i) .$$

A special case can be written when there are only two sets. In this case, write E_1 simply as E and E_2 as $\bar{E}$.

Then the law of total probability simplifies to

$$\Pr(A) = \Pr(A \cap E)\Pr(E) + \Pr(A \cap \bar{E})\Pr(\bar{E})$$

for any event A. This formula is the basis for event trees, which are frequently used to diagram the possibilities in an accident sequence.

The concepts of mutually exclusive events and statistically independent events are often confused. If E_1 and E_2 are mutually exclusive events and $\Pr(E_1)$ and $\Pr(E_2)$ are nonzero, $\Pr(E_1 \cap E_2) = \Pr(\emptyset) = 0$. From Equation A.1, $\Pr(E_2 \mid E_1) = 0$, which does not equal $\Pr(E_2)$. Thus, the two events are not independent. Mutually exclusive events cannot be independent and independent events cannot be mutually exclusive.

Equation A.2 can be used to calculate the probability of the intersection of a set of events (the probability that all the events occur simultaneously). For two events E_1 and E_2, the probability of simultaneous occurrence of the events is equal to the probability of E_1 times the probability of E_2 given that E_1 has already occurred. In general, the probability of the simultaneous occurrence of n events can be written as:

$$\Pr(E_1 \cap E_2 \cap \ldots \cap E_n) =$$

$$\Pr(E_1) \Pr(E_2 \mid E_1) \Pr(E_3 \mid E_2 \cap E_1) \ldots \Pr(E_n \mid E_{n-1} \cap \ldots \cap E_1),$$

which is referred to as the **chain rule**. This rule can be used to calculate the probability that a given accident sequence occurs, with E_1 denoting the initiating event and the remaining events corresponding to the failure or success of the systems that must function in order to mitigate such an accident.

The probability of occurrence of at least one of a set of statistically independent events yields a result that is important to PRA and fault tree applications. If E_1, E_2, ..., E_n are statistically independent events, the probability that at least one of the n events occurs is:

$$\Pr(E_1 \cup E_2 \cup \ldots \cup E_n) = \qquad \text{(A.3)}$$

$$1 - [1 - (\Pr(E_1)][1 - (\Pr(E_2)] \ldots [1 - (\Pr(E_n)],$$

which is equivalent (with expansion) to using the inclusion-exclusion rule. For the simple case where $\Pr(E_1) = \Pr(E_2) = \ldots = \Pr(E_n) = p$, the right-hand side of this expression reduces to $1 - (1 - p)^n$.

The general result in Equation A.3 has application in PRA and fault tree analysis. For example, for a system in which system failure occurs if any one of n independent events occurs, the probability of system failure is given by Equation A.3. These events could be failures of critical system components. In general, the events represent the modes by which system failure (the top event of the fault tree) can occur. These modes are referred to as the minimal cut sets of the fault tree and, if independent of each other (no minimal cut sets have common component failures), Equation A.3 applies. [See Vesely et al. (1981) for further discussion of fault trees and minimal cut sets.]

If the n events are <u>not</u> independent, the right side of Equation A.3 may be greater than or less than the left side. However, for an important situation that frequently arises in PRA, the right side of Equation A.3 forms an upper bound for the left side.

If the n events are cut sets that are positively associated [see Esary and Proschan (1970, 1963)], then the right side is an upper bound for $\Pr(E_1 \cup E_2 \cup \ldots \cup E_n)$ and is known as the **min cut upper bound** (NRC 1994). This name arises from common PRA applications where E_i is the i^{th} minimal cut set of a system or accident sequence of interest. In this case, the min cut upper bound is superior to the rare event approximation and can never exceed unity (as can happen with the rare event approximation). If the n events satisfy conditions similar to those of the rare event approximation, the min cut set upper bound is a useful approximation to the left side of Equation A.3. Note that the min cut upper bound is <u>not</u> applicable for mutually exclusive events.

A.4 Random Variables and Probability Distributions

A.4.1 Random Variables

A **random variable** is any rule that associates real numbers with the outcomes of an experiment. For example, the number of initiating events in one year, the number of failures to start in 12 demands, and the time to complete a repair of a pump are all random variables.

If the numbers associated with the outcomes of an experiment are all distinct and countable, the corresponding random variable is called a **discrete** random variable. The number of initiating events and the number of failures to start are examples of discrete random variables.

If the sample space contains an infinite number of outcomes (like those contained in any interval), the random variable is **continuous**. Time T is a common continuous random variable, for example, time to failure, time between failures, or time to repair, where the random variable T can assume any value over the range 0 to 4.

A.4.2 Probability Distributions

A **probability function** (introduced at the beginning of Section A.3) associates a probability with each possible value of a random variable and, thus, describes the distribution of probability for the random variable. For a discrete random variable, this function is referred to as a **discrete probability distribution function** (p.d.f.). A discrete p.d.f., commonly denoted by f, is also referred to as a **discrete distribution**, or **discrete probability mass function**.

If x denotes a value that the discrete random variable X can assume, the probability distribution function is often denoted $\Pr(x)$. The notation used here is that a random variable is denoted by an *upper-case letter* and an observed or observable value of the random variable (a number) is denoted by a *lower-case letter*. The sum of the probabilities over all the possible values of x must be 1. Thus, we write $f(x) = \Pr(X = x)$, and require $\sum_i f(x_i) = 1$.

Certain discrete random variables have wide application and have therefore been defined and given specific names. The two most commonly used discrete random variables in PRA applications are the **binomial** and **Poisson** random variables, which are presented in Section A.6.

A continuously distributed random variable has a **density function**, a nonnegative integrable function, with the area between the graph of the function and the horizontal axis equal to 1. This density function is also referred to as the **continuous probability distribution function** (p.d.f.). If x denotes a value that the continuous random variable X can assume, the p.d.f. is often denoted as $f(x)$. The probability that X takes a value in a region A is the integral of $f(x)$ over A. In particular,

$$\Pr(a \le X \le b) = \int_a^b f(x)dx$$

and

$$\Pr(x \# X \# x + \Delta x) \;.\; f(x)\,\Delta x \qquad \text{(A.4)}$$

for small Δx. Also,

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

The most commonly used continuous distributions in PRA are the **lognormal**, **exponential**, **gamma**, and **beta** distributions. Section A.7 summarizes the essential facts about these distributions, and also about less common but occasionally required distributions: **uniform**, **normal**, **Weibull**, **chi-squared**, **inverted gamma**, **logistic-normal**, **Student's *t***, ***F***, and **Dirichlet**.

A.4.3 Cumulative Distribution Functions

Discrete probability distributions provide point probabilities for discrete random variables and continuous p.d.f.s provide point densities for continuous random variables. A related function useful in probability and PRA is the **cumulative distribution function** (c.d.f.). This function is defined as the probability that the random variable assumes values less than or equal to the specific value x, and is denoted $F(x)$.

For a discrete random variable X, with outcomes x_i, and the corresponding probabilities $\Pr(x_i)$, $F(x)$ is the sum of the probabilities of all $x_i \# x$. That is,

$$F(x) \;=\; \Pr(X \le x) = \sum_{x_i \le x} \Pr(x_i)\;.$$

For a continuous random variable X, $F(x)$ is the area beneath the p.d.f. $f(x)$ up to x. That is, $F(x)$ is the integral of $f(x)$:

$$F(x) = \Pr(X \# x) = \int_{-\infty}^{x} f(y)dy\;.$$

If X takes on only positive values, the lower limit of integration can be set to 0. The upper limit is x, and $f(x)$ is the derivative of $F(x)$. Note that, because $F(x)$ is a probability, $0 \# F(x) \# 1$. If X ranges from -4 to +4, then

$F(-4) = 0$ and $F(+4) = 1$.

If X has a restricted range, with a and b being the lower and upper limits of X respectively, $a < X < b$, then

$F(a) = 0$ and $F(b) = 1$.

Also, $F(x)$ is a nondecreasing function of x, that is,

if $x_2 > x_1$, $F(x_2) \$ F(x_1)$.

Another important property of $F(x)$ is that

$\Pr(x_1 < X \# x_2) = F(x_2) \; ! \; F(x_1)$

for discrete random variables and

$\Pr(x_1 \# X \# x_2) = F(x_2) \; ! \; F(x_1)$

for continuous random variables.

An example of a p.d.f. and the associated c.d.f. for a continuous distribution is shown in Figure A.2.

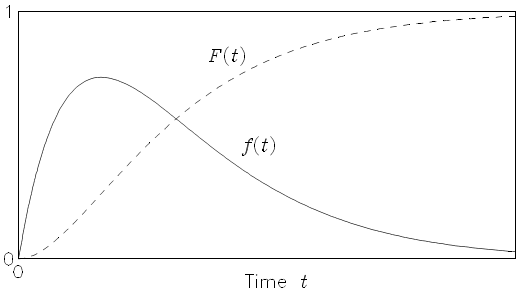


Figure A.2 Probability density function (p.d.f.) and cumulative distribution function (c.d.f.).

A.4.4 Reliability and Hazard Functions

A.4.4.1 Definitions

There are also characterizations that have special interpretations for time-to-failure distributions. Let T denote the random time to failure of a system. The **reliability function** of a system is defined as

$R(t) = \Pr(T > t)$.

Hence, $R(t)$, called the **reliability** at time t, is the probability that the system does not fail in the time interval $(0, t)$ or equivalently, the probability that the system is still operating at time t. (This discussion uses the notation (a, b) to mean the set of times $> a$ and $\# b$, but the distinction between $<$ and $\#$ is a mathematical fine point, not important in practice.) The reliability function is also sometimes called the **survival function**. It is equal to $1 \; ! \; F(t)$.

When used as a reliability criterion, it is common to state a time, say t_0, called the **mission time**, and require for a system that the reliability at mission time t_0 be at least some prescribed level, say R_0. For example, a pump might be required to operate successfully for at least 12 hours with probability at least 0.95. The requirement in this case is $R_0 = 0.95$ and $t_0 = 12$. In terms of the reliability function, this would mean $R(12) \$ 0.95$. One interpretation would be that such a pump would perform for the required mission time for 95% of the situations when it is called on to do so. Another interpretation is that 95% of all such pumps would perform as required.

Consider a system that operates for a particular mission time, unless it fails. If it fails, no immediate repairs are attempted, so some authors call the system **nonrepairable**. A common way to characterize this system's reliability is in terms of the **hazard function**. Suppose that the system is still operating at time t, and consider the probability that it will fail in a small interval of time $(t, t + \Delta t)$. This is the conditional probability $\Pr(t < T \# t + \Delta t \; * \; T > t)$. The hazard function, h, is defined so that when Δt is small,

$$h(t)\,\Delta t \;.\; \Pr(t < T \# t + \Delta t \; * \; T > t) \;. \qquad \text{(A.5)}$$

This function is also encountered, under the name of λ, in some treatments of Poisson processes. Equation A.5 gives, approximately,

$$h(t)\Delta t \approx \frac{\Pr(t < T \le t + \Delta t)}{\Pr(T > t)} \approx \frac{f(t)\Delta t}{R(t)}$$

This is the basis for the formal definition of h:

$$h(t) = \frac{f(t)}{R(t)}$$

For details, see Bain and Engelhardt (1992, p. 541). Equation A.5 is analogous to Equation A.4, except that the probability in Equation A.5 is conditional on the system having survived until t, whereas Equation A.4 refers to all systems in the original population, either still surviving or not. Suppose a large number, say N, of identical systems are put into operation at time $t = 0$, and n is the number which fail in the interval $(t, t + \Delta t)$. It follows that $f(t)\Delta t \approx n/N$, the observed relative frequency of systems failed in the interval $(t, t + \Delta t)$. On the other hand, if N_t denotes the number of the original N systems which are still in operation at time t, then $h(t)\Delta t \approx n/N_t$, the observed relative frequency of *surviving* systems which fail in this same interval. Thus, $f(t)$ is a measure of the risk of failing at time t for any system in the original set, whereas $h(t)$ is a measure of the risk of failing at time t, but only for systems that have survived this long.

The hazard function is used as a measure of "aging" for systems in the population. If $h(t)$ is an increasing function, then systems are aging or wearing out with time. Of course, in general the hazard function can exhibit many types of behavior other than increasing with time. In actuarial science the hazard function is called the **force of mortality**, and it is used as a measure of aging for individuals in a population. More generally, the hazard function gives an indication of "proneness to failure" of a system after time t has elapsed. Other terms which are also used instead of hazard function are **hazard rate** and **failure rate**. The term *failure rate* is often used in other ways in the literature of reliability [see Ascher and Feingold (1984), p. 19].

A.4.4.2 Relations among p.d.f., Reliability, and Hazard

Any one of the functions F, f, R, and h completely characterizes the distribution, and uniquely determines the other three functions. The definition

$$h(t) = \frac{f(t)}{R(t)}$$

was given above. The right side can be written as the derivative of $-\ln[R(t)]$, leading to

$$R(t) = \exp\left(-\int_0^t h(u)du\right) = \exp\left(-H(t)\right)$$

where the function $H(t)$ is called the **cumulative hazard function**. The reliability function, $R(t)$, and the c.d.f., $F(t) = 1 - R(t)$, are therefore uniquely determined by the hazard function, $h(t)$, and the p.d.f. can be expressed as

$$f(t) = h(t)\exp\left(-\int_0^t h(u)du\right).$$

Figure A.3 shows the reliability, hazard and the cumulative hazard function for the example of Figure A.2.

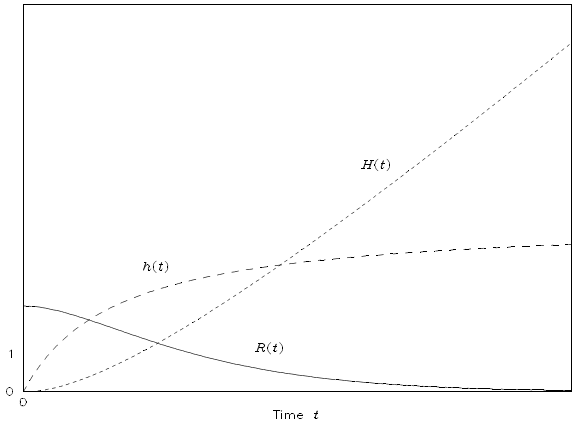


Figure A.3 The reliability function, hazard function and cumulative hazard function.

The hazard function in Figure A.3 is an increasing function of time. Therefore, it would be consistent with systems with a dominant wear-out effect for the entire life of the system. The lifetime of a system may be divided into three typical intervals: the **burn-in** or **infant** period, the **random** or **chance** failure period, and the **wear-out** period. During the useful period, the dominant cause of failures is "random" failures. For example, systems might fail due to external causes such as power surges or other environmental factors rather than problems attributable to the defects or wear-out in the systems. This example is somewhat idealized because for many types of systems the hazard function will tend to increase slowly during the later stages of the chance failure period. This is particularly true of mechanical systems. On the other hand, for many electrical components such as transistors and other solid-state devices, the hazard function remains fairly flat once the burn-in failure period is over.

A.4.5 Joint, Marginal, and Conditional Distributions

Many statistical methods are based on selecting a sample of size n from a probability distribution $f(x)$. Such a sample is denoted by

$$(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = (x_1, x_2, ..., x_n),$$

where $x_1, x_2, ..., x_n$ are the actual values of the random variable X which has the distribution $f(x)$.

The concepts of simultaneous events and joint, marginal, and conditional probability, discussed in Section A.3, also pertain to random variables and probability distributions. Two random variables X_1 and X_2 (both continuous, both discrete, or one of each) can have a **joint distribution**, or joint p.d.f., denoted $f(x_1, x_2)$. The point (x_1, x_2) can be thought of as a point in two-dimensional Euclidean space. Similarly, n random variables have joint distribution $f(x_1, x_2, ..., x_n)$. Also, the n random variables have joint cumulative distribution $F(x_1, x_2, ..., x_n)$.

The **marginal distribution** of X_i is defined as the joint p.d.f. integrated (for continuous random variables) or summed (for discrete random variables) over the n ! 1 other corresponding dimensions, resulting in a function of x_i alone. Thus, the marginal distribution of X_i is the unconditional p.d.f. of X_i, $f_i(x_i)$.

The **conditional distribution** of X_1 given X_2, denoted $f(x_1 \mid x_2)$, is defined by

$$f(x_1 \mid x_2) = \frac{f(x_1, x_2)}{f_2(x_2)},$$

where $f_2(x_2)$ ... 0. This conditional distribution can be shown to satisfy the requirements of a probability function. Sampling from a conditional p.d.f. would produce only those values of X_1 that could occur for a given value of $X_2 = x_2$. The concept of a conditional distribution also extends to n random variables.

Two random variables X_1 and X_2 are independent if their joint p.d.f. is equal to the product of the two individual p.d.f.s. That is,

$f(x_1, x_2) = f(x_1)\, f(x_2)$.

In general, $X_1, X_2, ..., X_n$ are **independent random variables** if

$f(x_1, x_2, ..., x_n) = f(x_1) \cdot f(x_2) \cdot \ldots \cdot f(x_n)$.

A.4.6 Characterizing Random Variables and Their Distributions

A.4.6.1 Distribution Characteristics

Probability distributions have many characteristics of interest, some of which are described by **distribution parameters**. The term parameter is used to refer to a fixed characteristic. In contrast to a statistic, which changes from sample to sample, a parameter for a particular distribution is a constant and does not change. However, when a parameter's value is not known, sample statistics can be used to estimate the parameter value. Parameter estimation is discussed in Appendix B.

A very useful distribution characteristic is the parameter that serves as a measure of central tendency, which can be viewed as a measure of the middle of a distribution. When a change in the parameter slides the distribution sideways, as with the mean of a normal distribution, the parameter is referred to as the **location parameter**. It serves to locate the distribution along the horizontal axis. Sometimes, however, a change in the parameter squeezes or stretches the distribution toward or away from zero, as with the mean of the exponential distribution. In that case, the parameter is a **scale parameter**.

In any case, the most common measure of central tendency is the **mean**, μ, of the distribution, which is a weighted average of the outcomes, with the weights being probabilities of outcomes. For a discrete random variable X,

$$\mu_X = \sum_i x_i \Pr(x_i).$$

For a continuous random variable X,

$$\mu_X = \int_{-\infty}^{\infty} x\, f(x)\, dx.$$

(In Section A.4.6.2 below, the mean of X will be denoted $E(X)$, the "expected value" of X.)

Another distribution characteristic commonly used as a measure of central tendency, or location, is the **median**. For a continuous distribution, the median is the point along the horizontal axis for which 50% of the area under the p.d.f. lies to its left and the other 50% to its right. The median of a random variable, X, is commonly designated med(X) or $x_{.50}$ and, for a continuous distribution, is the value for which $\Pr(X \# x_{.50}) = .50$ and $\Pr(X \$ x_{.50}) = .50$. In terms of the cumulative distribution, $F(x_{.50}) = .50$. The median is a specific case of the general 100 ″th **percentile**, $x_{\text{″}}$, for which $F(x_{\text{″}}) =$ ″. When the factor of 100 is dropped, $x_{\text{″}}$ is called the ″**quantile**. Along with the median as the 50th percentile (or equivalently, the 0.5 quantile), the 25th and 75th percentiles are referred to as **quartiles** of a distribution.

Figure A.4 shows the quartiles, $x_{0.25}$ and $x_{0.75}$, the median, $x_{0.50}$, and the mean. The quartiles and the median divide the area under the density curve into four pieces, each with the same area. Note that the mean is greater than the median in this example, which is the usual relation when the density has a long right tail, as this one does.

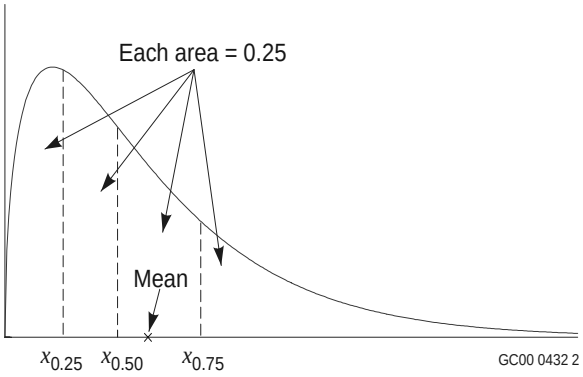


Figure A.4 Density, showing quartiles, median, and mean.

Figure A.5 shows the same quantities plotted with the c.d.f. By definition, the q quantile, x_q, satisfies $F(x_q) = q$.

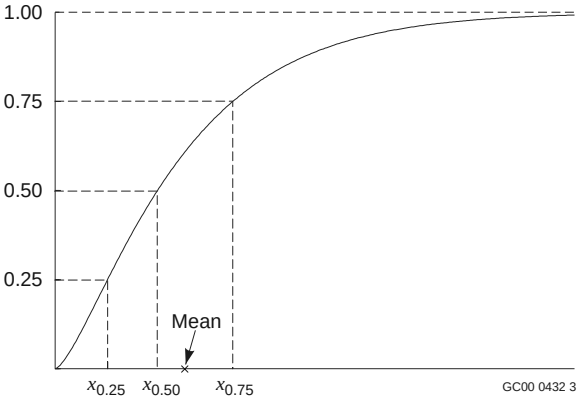


Figure A.5 Cumulative distribution function (c.d.f.) showing quartiles, median, and mean.

The mean and the median are used to measure the center or location of a distribution. Since the median is less affected by tail-area probabilities, it can be viewed as a better measure of location than the mean for highly-skewed distributions. For symmetric distributions, the mean and median are equivalent.

A different measure of center or location of a distribution is the **mode**, which indicates the most probable outcome of a distribution. The mode is the point along the horizontal axis where the "peak" or maximum of the p.d.f. is located. Note that the mode does not necessarily have to be near the middle of the distribution. It simply indicates the most likely value of a distribution. Note also that a peak does not have to exist and, in some cases, more than one peak can exist.

Another important characteristic of a distribution is its **variance**, denoted by σ^2. The variance is the average of the squared deviations from the mean. The **standard deviation**, σ, of a distribution is the square root of its variance. Both the variance and standard deviation are measures of a distribution's spread or dispersion. For a discrete random variable X,

$$\sigma_X^2 = \sum_i (x_i - \mu)^2 \Pr(x_i) .$$

For a continuous random variable X,

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx .$$

Though less used than the mean and variance, the **skewness** is defined as

$E(X - \mu)^3/\sigma^3$.

It measures asymmetry. It is usually positive if the density has a longer right tail than left tail, and negative if the density has a longer left tail than right tail. For example, the density in Figure A.4 has positive skewness.

A.4.6.2 Mathematical Expectation

The definitions of distribution means and variances arise from **mathematical expectation** and **moments of a distribution**, which form an important method for calculating the parameters of a known p.d.f. In general, the **expectation** (**expected value** or mathematical expectation) of a function $g(X)$, denoted $E[g(X)]$, is

$$E[g(X)] = \sum_i g(x_i) \Pr(x_i) ,$$

when X is discrete, and

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx ,$$

when X is continuous.

Because of their wide use, several expectations have special names. For $g(X) = X$, the expectation $E(X)$ becomes the mean of X. Thus, the mean is also commonly referred to as the expected value (or expectation) of the random variable X. In addition, for $g(X) = X$, the expectation $E(X)$ is known as the **first moment about the origin**.

The variance, σ_X^2, also denoted by Var(X), of a distribution is defined by mathematical expectation with $g(X) = (X - \mu_X)^2$. Thus,

$$\text{Var}(X) = \sigma_X^2 = E[(X - \mu_X)^2] = E(X^2) - [E(X)]^2,$$

which is known as the **second moment about the mean**.

Ordinary moments (moments about the origin) of a random variable X are defined as

$$M_r = E(X^r),$$

for $r = 1, 2, \ldots$. Thus,

$$\text{Var}(X) = \sigma_X^2 = E(X^2) - [E(X)]^2 = M_2 - M_1^2.$$

Central moments (moments about the mean) of a random variable X are defined as being equal to $E[(X - \mu)^r]$ for $r = 2, 3, \ldots$. The ordinary and central moments can be seen to define characteristics of distributions of random variables.

An important rule of expectation commonly used in PRA is that the expected value of a product of independent random variables is the product of their respective expected values. That is, $E(X_1 \cdot X_2 \cdot \ldots \cdot X_n) = E(X_1) \cdot E(X_2) \cdot \ldots \cdot E(X_n)$ when all X_i are independent. This rule also applies to conditionally independent random variables. If the random variables $X_2, X_3, \ldots, X_n$ are all conditionally independent given $X_1 = x_1$, then

$$f(x_2, x_3, \ldots, x_n | x_1) = f(x_2 | x_1) \cdot f(x_3 | x_1) \cdot \ldots \cdot f(x_n | x_1).$$

It follows that

$$E(X_2 \cdot X_3 \cdot \ldots \cdot X_n | x_1) = E(X_2 | x_1) \cdot E(X_3 | x_1) \cdot \ldots \cdot E(X_n | x_1).$$

Thus,

$$E(X_1 \cdot X_2 \cdot \ldots \cdot X_n) = E[X_1 \cdot E(X_2 | x_1) \cdot E(X_3 | x_1) \cdot \ldots \cdot E(X_n | x_1)].$$

The following facts are also often useful:

- $E(\sum_i X_i) = \sum_i E(X_i)$, whether or not the X_is are independent.
- $\text{Var}(\sum_i X_i) = \sum_i \text{Var}(X_i)$, if the X_is are independent.
- $E(aX + b) = aE(X) + b$.
- $\text{Var}(aX + b) = a^2\text{Var}(X)$.
- The last two give useful special cases when $a = 1$ or $b = 0$.

A.4.6.3 Moment-Generating Functions

Another special mathematical expectation is the **moment-generating function** of a random variable. For a random variable X with p.d.f. $f(x)$, the moment-generating function of X (or of the distribution) is defined by $M(t) = E(e^{tX})$, if M exists for some interval $-h < t < h$. Therefore, if X is a continuous random variable,

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx .$$

If X is a discrete random variable,

$$M(t) = \sum_i e^{tx_i} f(x_i) .$$

Note that not every distribution has a moment-generating function.

The importance of the moment-generating function is that, when it does exist, it is unique and completely specifies the distribution of the random variable. If two random variables have the same moment-generating function, they have the same distribution.

It can be shown that the moments of a distribution can be found from the series expansion of $M(t)$. The moments of the distribution can also be determined from the moment-generating function by differentiating the moment-generating function with respect to t and setting $t = 0$. See Martz and Waller (1991) and any of several mathematical statistics texts, such as Hogg and Craig (1995), for further details on moment-generating functions.

A.4.6.4 Covariance and Correlation

For two random variables, X and Y, with means μ_x and μ_y, the expected value $E[(X - \mu_x)(Y - \mu_y)]$ is called the **covariance** of X and Y, denoted Cov(X, Y). The covariance of X and Y divided by the product of the standard deviations of X and Y is called the **correlation coefficient** (or correlation) between X and Y, denoted Cor(X, Y). That is,

$$\mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

$$= \frac{E(X-\mu_X)E(Y-\mu_Y)}{\sqrt{E[(X-\mu_X)^2]E[(Y-\mu_Y)^2]}} \ .$$

The correlation coefficient measures the degree of association between X and Y, that is, the strength of a linear relationship between X and Y. It is always between !1 and 1. Positive correlation (correlation > 0) means that X and Y tend to be large together and small together in a linear way. Negative correlation means that X tends to be large when Y is small and vice versa, in a linear way. If X and Y are independent, then their correlation is zero. The converse is not true; examples can be constructed where X and Y are dependent (in a nonlinear way) yet have zero correlation.

A.4.7 Distribution of a Transformed Random Variable

This section considers the distribution of $Y = h(X)$, when X has a known distribution and h is a known function. The problem is straightforward when X has a discrete distribution. When X is continuous and h is monotone, either increasing or decreasing, the c.d.f.s are also related in the natural way, as follows. Let F be the c.d.f. of X and let G be the c.d.f. of Y. Then we have

$$G(y) = \Pr(Y \# y) = \Pr[h(X) \# y] \ .$$

If h is monotone increasing, this equals

$$\Pr[X \# h^{-1}(y)] = F(x),$$

where x and y are related by $y = h(x)$, $x = h^{-1}(y)$. In summary, $G(y) = F(x)$.

If, instead, h is monotone decreasing, then a similar argument gives

$$G(y) = 1 \ ! \ F(x) \ .$$

The surprise comes with the densities. Differentiate both sides of either of the above equations with respect to y, to obtain the density of y. This involves using the chain rule for differentiation. The result is

$$g(y) = f(x)\left|\frac{dx}{dy}\right| \ .$$

That is, the density of Y is not simply equal to the density of X with a different argument. There is also a multiplier, the absolute value of the derivative.

Two important special cases are given here. If $Y = \exp(X)$, then

$$g(y) = f[\ln(y)](1/y) \ .$$

If $Y = 1/X$, then

$$g(y) = f(1/y)(1/y^2) \ .$$

These formulas form the basis for the densities of the lognormal distribution and the inverted gamma distribution.

A.5 Bayes' Theorem

It is frequently desired to calculate the probability of an event A given that another event B has occurred at some prior point in time. It can also be of interest to calculate the probability that a state of nature exists given that a certain sample is observed, for example, belonging to a certain population based on a sample measurement or observation. Conditional probability leads directly to **Bayes' Theorem**, which, along with subjective probability, forms the basis for Bayesian inference commonly used in PRA.

Recall the definition of a partition from Section A.3: $A_1, A_2, ..., A_n$ are a partition of the sample space if the are mutually exclusive and their union equals the entire sample space. Bayes' Theorem states that if $A_1, A_2, ..., A_n$ are a partition of the sample space and if B is any other event such that $\Pr(B) > 0$, then

$$\Pr(A_i|B) = \frac{\Pr(B|A_i)\ \Pr(A_i)}{\Pr(B)} \ , \qquad \text{(A.6)}$$

where

$$\Pr(B) = \sum_{j=1}^{n} \Pr(B|A_j)\Pr(A_j) \ .$$

This last equation is the law of total probability (Section A.3). Equation A.6 follows from the definition of conditional probability in Equation A.1:

$$\Pr(A_i|B) = \frac{\Pr(B \cap A_i)}{\Pr(B)} = \frac{\Pr(B|A_i)\ \Pr(A_i)}{\Pr(B)} \ .$$

The Pr(A_i*B) is the **posterior** (or a posteriori) probability for the event A_i, meaning the probability of A_i once B is known. The Pr(A_i) is the **prior** (or a priori) probability of the event A_i before experimentation or observation. The event B is the observation. The Pr(B*A_i) is the probability of the observation given that A_i is true. The denominator serves as a normalizing constant.

Calculating the posterior probabilities Pr(A_i*B) requires knowledge of the probabilities Pr(A_i) and Pr(B*A_i), i = 1, 2, ..., n. The probability of an event can often be determined if the population is known, thus, the Pr(B*A_i) can be determined. However, the Pr(A_i), i = 1, 2, ..., n, are the probabilities that certain states of nature exist and are either unknown or difficult to ascertain. These probabilities, Pr(A_i), are called prior probabilities for the events A_i because they specify the distribution of the states of nature prior to conducting the experiment.

Application of Bayes' Theorem utilizes the fact that Pr(B*A_i) is easier to calculate than Pr(A_i*B). If probability is viewed as degree of belief, then the prior belief is changed, by the test evidence, to a posterior degree of belief. In many situations, some knowledge of the prior probabilities for the events A_1, A_2, ..., A_n exists. Using this prior information, inferring which of the sets A_1, A_2, ..., A_n is the true population can be achieved by calculating the Pr(A_i*B) and selecting the population that produces the highest probability.

Equation A.6 pertains to disjoint discrete events and discrete probability distributions. Bayes' Theorem has analogous results for continuous p.d.f.'s. The continuous version is given here. Suppose X is a discrete or continuous random variable, with p.d.f. depending on parameter θ, and with conditional p.d.f. of X, given θ, specified by $f(x|\theta)$. Suppose that θ has a continuous probability distribution with p.d.f. $g(\theta)$. This can happen in two ways: either θ is a possible value of the random variable Θ (using the convention of denoting random variables with uppercase letters), or else θ is an uncertain parameter with a subjective uncertainty distribution. The second case is the more common one. Call $g(\theta)$ the **prior p.d.f.** Then for every x such that $f(x) > 0$ exists, the **posterior p.d.f.** of θ, given $X = x$, is

$$g(\theta|x) = \frac{f(x|\theta)g(\theta)}{f(x)}, \qquad \text{(A.7)}$$

where

$$f(x) = \int f(x|\theta)g(\theta)d\theta$$

is the marginal p.d.f. of X. Again, the prior and posterior p.d.f.'s can be used to represent the probability of various values θ prior to and posterior to observing a value of another random variable X. This is valid whether "probability of θ" has a frequentist or subjective interpretation.

A.6 Discrete Random Variables

A.6.1 The Binomial Distribution

The **binomial** distribution describes the number of failures X in n independent trials. The random variable X has a binomial distribution if:

1. The number of random trials is one or more and is known in advance.
2. Each trial results in one of two outcomes, usually called success and failure (although they could be pass-fail, hit-miss, defective-nondefective, etc.).
3. The outcomes for different trials are statistically independent.
4. The probability of failure, p, is constant across trials.

Equal to the number of failures in the n trials, a binomial random variable X can take on any integer value from 0 to n. The probability associated with each of these possible outcomes, x, is defined by the binomial(n, p) p.d.f. as

$$\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x},$$

$$x = 0, \ldots, n.$$

Here

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

is the **binomial coefficient**. The symbol

$$n! = n(n-1)(n-2) \ldots (2)(1)$$

denotes n factorial, with 0! defined to be equal to 1. This binomial coefficient provides the number of ways that exactly x failures can occur in n trials (number of combinations of n trials selected x at a time).

The binomial distribution has two parameters, n and p, of which n is known. (Although n may not always be known exactly, it is treated as known in this handbook.)

The mean and variance of a binomial(n, p) random variable X are

$E(X) = np$

and

$\mathrm{Var}(X) = np(1 - p)$.

Figure A.6 shows three binomial probability distribution functions, with parameter $p = 0.25$, and $n = 4$, 12, and 40. In each case, the mean is np. The means have been aligned in the three plots.

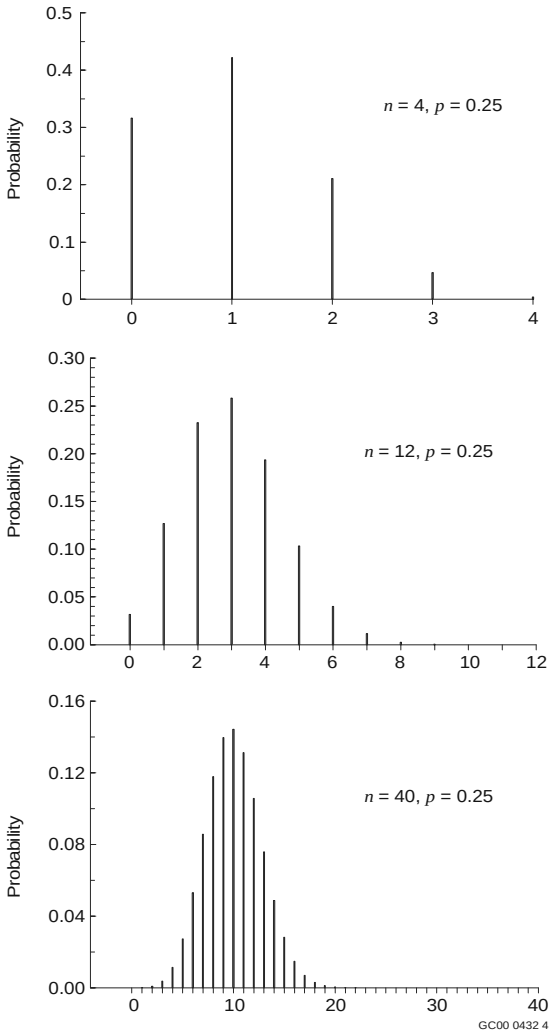


Figure A.6 Three binomial probability distribution functions.

A.6.2 The Poisson Distribution

The **Poisson** distribution provides a discrete probability model that is appropriate for many random phenomena that involve counts. Examples are counts per fixed time interval of the number of items that fail, the number of customers arriving for service, and the number of telephone calls occurring. A common use of the Poisson distribution is to describe the behavior of many rare event occurrences. The Poisson distribution is also frequently used in applications to describe the occurrence of system or component failures under steady-state conditions.

The count phenomena that occur as Poisson random variables are not necessarily restricted to occurring over a time interval. They could also be counts of things occurring in some region, such as defects on a surface or within a certain material. A process that leads to a Poisson random variable is said to be a **Poisson process**.

The Poisson distribution describes the total number of events occurring in some interval of time t (or space). The p.d.f. of a Poisson random variable X, with parameter $\nu = \theta t$, is

$$\Pr(X = x) = \frac{e^{-\mu}\mu^{x}}{x!} = \frac{e^{-\lambda t}(\lambda t)^{x}}{x!}, \qquad \text{(A.8)}$$

for $x = 0, 1, 2, \ldots$, and $x! = x(x - 1)(x - 2) \ldots (2)(1)$, as defined previously.

The Poisson distribution has a single parameter ν, denoted Poisson(ν). If X denotes the number of events that occur during some time period of length t, then X is often assumed to have a Poisson distribution with parameter $\nu = \theta t$. In this case, X is considered to be a Poisson process with **intensity** $\theta > 0$ (Martz and Waller 1991). The variable θ is also referred to as the **event rate** (or **failure rate** when the events are failures). Note that θ has units 1/time; thus, $\theta t = \nu$ is dimensionless.

If only the total number of occurrences for a single time period t is of interest, the form of the p.d.f. in Equation A.8 using ν is simpler. If the event rate, θ, or various time periods, t, are of interest, the form of the p.d.f. in Equation A.8 using θt is more useful.

The expected number of events occurring in the interval 0 to t is $\nu = \theta t$. Thus, the mean of the Poisson distribution is equal to the parameter of the distribution, which is why ν is often used to represent the parameter. The variance of the Poisson distribution is also equal to the parameter of the distribution. Therefore, for a Poisson(ν) random variable X,

$E(X) = \mathrm{Var}(X) = \nu = \theta t$.

Figure A.7 shows three Poisson probability distribution functions, with means λt = 1.0, 3.0, and 10.0, respectively. The three means have been aligned in the graphs. Note the similarity between the Poisson distribution and the binomial distribution when $\lambda t = np$ and n is not too small.

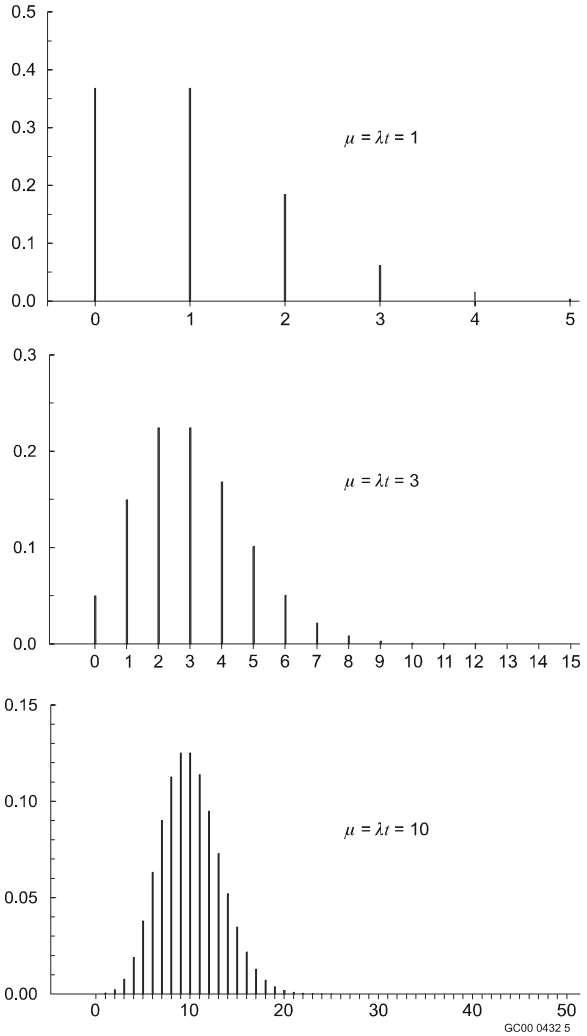


Figure A.7 Three Poisson probability distribution functions.

Several conditions are assumed to hold for a Poisson process that produces a Poisson random variable:

1. For small intervals, the probability of exactly one occurrence is approximately proportional to the length of the interval (where δ, the event rate or intensity, is the constant of proportionality).
2. For small intervals, the probability of more than one occurrence is essentially equal to zero (see below).
3. The numbers of occurrences in two non-overlapping intervals are statistically independent.

More precise versions of condition 2 are: (1) the probability of more than one event occurring in a very short time interval is negligible in comparison to the probability that only one event occurs (Meyer 1970), (2) the probability of more than one event occurring in a very short time interval goes to zero faster than the length of the interval (Pfeiffer and Schum 1973), and (3) simultaneous events occur only with probability zero (Çinlar 1975). All of these versions have the practical interpretation that common cause events do not occur. The phrase "do not occur" is used in this handbook, as it is in Thompson (1981).

The Poisson distribution also can serve as an **approximation to the binomial distribution**. Poisson random variables can be viewed as resulting from an experiment involving a large number of trials, n, that each have a small probability of occurrence, p, of an event. However, the rare occurrence is offset by the large number of trials. As stated above, the binomial distribution gives the probability that an occurrence will take place exactly x times in n trials. If $p = \lambda t/n$ (so that p is small for large n), and n is large, the binomial probability that the rare occurrence will take place exactly x times is closely approximated by the Poisson distribution with $\lambda t = np$. In general, the approximation is good for large n, small p, and moderate λt (say $\lambda t \# 20$) [see Derman et al. (1973)].

The Poisson distribution is important because it describes the behavior of many rare event occurrences, regardless of their underlying physical process. It also has many applications to describing the occurrences of system and component failures under steady-state conditions. These applications utilize the relationship between the Poisson and **exponential** (continuous random variable, see Section A.7.4) distributions: the times between successive events follow an exponential distribution.

A.7 Continuous Random Variables

A.7.1 The Uniform Distribution

A **uniform** distribution, also referred to as a rectangular distribution, represents the situation where any value in a specified interval, say [a, b], is equally likely. For a uniform random variable, X, because the outcomes are equally likely, $f(x)$ is equal to a constant. The p.d.f. of a uniform distribution with parameters a and b, denoted uniform(a, b), is

$$f(x) = \frac{1}{b-a}$$

for $a \# x \# b$.

Figure A.8 shows the density of the uniform(a, b) distribution.

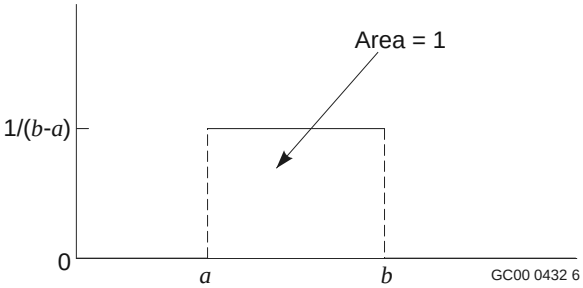


Figure A.8 Density of uniform(a, b) distribution.

The mean and variance of a uniform(a, b) distribution are

$$E(X) = \frac{b+a}{2}$$

and

$$\mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

A.7.2 The Normal Distribution

One of the most widely encountered continuous probability distributions is the **normal** distribution, which has the familiar bell shape and is symmetrical about its mean value. The importance of the normal distribution is due to: (1) its applicability in describing a very large number of random variables that occur in nature and (2) the fact that certain useful functions of nonnormal random variables are approximately normal. Details on the derivation of the normal distribution can be found in many basic mathematical statistics textbooks [e.g., Hogg and Craig (1995)].

The normal distribution is characterized by two parameters, μ and σ. For a random variable, X, that is normally distributed with parameters μ and σ, the p.d.f. of X is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \qquad \text{(A.9)}$$

for $-\infty < x < \infty$, $-\infty < \mu < \infty$, and $\sigma > 0$. Increasing μ moves the density curve to the right and increasing σ spreads the density curve out to the right and left while lowering the peak of the curve. The units of μ and σ are the same as for X.

The mean and variance of a normal distribution with parameters μ and σ are

$E(X) = \mu$

and

$\mathrm{Var}(X) = \sigma^2$.

The normal distribution is denoted normal(μ, σ^2).

Figure A.9 shows two normal(μ, σ^2) densities. The distribution is largest at μ and is more concentrated around μ when σ is small than when σ is large.

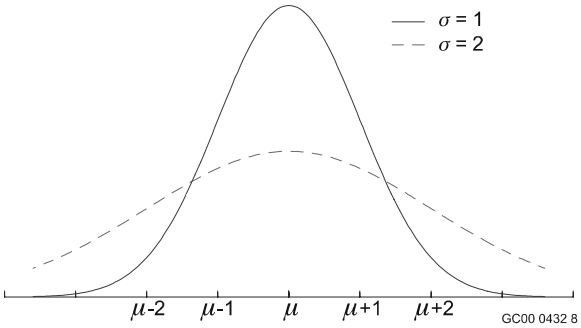


Figure A.9 Two normal densities.

Note the similarity of the normal density to a binomial p.d.f. with large np or a Poisson p.d.f. with large λ. This illustrates the fact that a normal distribution can sometimes be used to approximate those distributions.

The normal(0, 1) distribution is called the **standard normal** distribution, which, from Equation A.9, has p.d.f.

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \qquad \text{(A.10)}$$

for $-\infty < x < \infty$. The cumulative distribution of the standard normal distribution is denoted by Φ. Tables for the standard normal distribution are presented in Appendix C and in almost all books on statistics.

It can be shown that the transformed random variable $Z = (X - \mu)/\sigma$ is normal(0, 1). Thus, to calculate probabilities for a normal(μ, σ^2) random variable, X, when $\mu \neq 0$ and/or $\sigma^2 \neq 1$, the tables for the standard normal can be used. Specifically, for any number a,

$$\Pr[X \le a] = \Pr[(X - \mu)/\sigma \le (a - \mu)/\sigma]$$

$$= \Pr[Z \le (a - \mu)/\sigma]$$

$$= M[(a - \mu)/\sigma] .$$

Part of the importance of the normal distribution is that it is the distribution that sample sums and sample means tend to possess as n becomes sufficiently large. This result is known as the **central limit theorem**, which states that, if X_1, X_2, ..., X_n, are independent random variables, each with mean μ and variance σ^2, the sum of these n random variables, $\sum_i X_i$, tends toward a normal($n\mu$, $n\sigma^2$) distribution for large enough n. Since the sample mean is a linear combination of this sum, the central limit theorem also applies. Thus, $\overline{X} = \sum_i X_i/n$ tends to a normal(μ, σ^2/n) distribution. The importance of the central limit theorem is it can be used to provide approximate probability information for the sample sums and sample means of random variables whose distributions are unknown. Further, because many natural phenomena consist of a sum of several random contributors, the normal distribution is used in many broad applications.

Because a binomial random variable is a sum, it tends to the normal distribution as n gets large. Thus, the normal distribution can be used as an **approximation to the binomial distribution**. One rule of thumb is that the approximation is adequate for $np \ge 5$.

A Poisson random variable also represents a sum and, as presented previously, can also be used as an approximation to the binomial distribution. It follows that the normal distribution can serve as an **approximation to the Poisson distribution** when $\mu = \delta t$ is large. One rule of thumb is that the approximation is adequate for $\mu \ge 5$.

A.7.3 The Lognormal Distribution

Use of the **lognormal distribution** has become increasingly widespread. It is commonly used as a distribution for failure time and in maintainability analysis (Martz and Waller 1991). It has also been widely used as a prior distribution for unknown positive parameters.

The lognormal distribution arises from the *product* of many independent random variables. If $Y = Y_1 \cdot Y_2 \cdot \ldots \cdot Y_n = \prod_i Y_i$ is the product of n independent positive random variables that are (nearly) identically distributed, then $\ln(Y) = \ln(\prod_i Y_i) = \sum_i \ln(Y_i)$ is a sum that tends toward a normal distribution.

The distribution of Y is defined to be lognormal when the distribution of $\ln(Y)$ is normal. That is, when Y is lognormal, $\ln(Y)$ is normal(μ, σ^2). The parameters of the lognormal distribution are μ and σ, the parameters from the underlying normal distribution. For a random variable, Y, that is lognormally distributed with parameters μ and σ, denoted lognormal(μ, σ^2), the p.d.f. of Y is

$$f(y) = \frac{1}{\sigma y \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2} (\ln y - \mu)^2 \right]$$

for $0 < y < \infty$, $-\infty < \mu < \infty$, and $\sigma > 0$. Note the y in the denominator, for reasons explained in Section A.4.7. The mean and variance of a lognormal(μ, σ^2) distribution are

$$E(Y) = \exp(\mu + \sigma^2/2)$$

and

$$\mathrm{Var}(Y) = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1].$$

In addition, the median of a lognormal distribution is $\exp(\mu)$ and the mode is $\exp(\mu - \sigma^2)$. See Martz and Waller (1991) for more information on the lognormal distribution.

Sometimes the median of $Y = \exp(\mu)$ is used as a parameter. In addition, a parameter commonly used in PRA is the **error factor** (EF), where $EF = \exp(1.645\sigma)$. This definition causes EF to satisfy

$$\Pr[\mathrm{med}(Y)/EF \le Y \le \mathrm{med}(Y) * EF] = 0.90.$$

Figure A.10 shows three lognormal densities. The value $\mu = -7$ corresponds to a median of about $1.E-3$. [More exactly, it corresponds to $\exp(-7) = 9.E-4$.] The value $\mu = -6.5$ corresponds to a median of about $1.5E-3$. The value $\sigma = 0.67$ corresponds to an error factor EF = 3, and $\sigma = 1.4$ corresponds to an error factor EF = 10.

The two distributions with $\sigma = 0.67$ and different values of μ have essentially the same shape, but with different scales. The larger μ corresponds to spreading the distribution out more from zero. The distribution with $\sigma = 1.4$, and therefore EF = 10, has a very skewed distribution.

To calculate probabilities for a lognormal(μ, σ^2) random variable, Y, the tables for the standard normal can be used. Specifically, for any number b,

$$\Pr[\, Y \le b \,] = \Pr[\, \ln(Y) \le \ln(b) \,]$$

$$= \Pr[\, X \le \ln(b) \,]$$

$$= \Phi[\, (\ln(b) - \mu)/\sigma \,],$$

where $X = \ln(Y)$ is normal(μ, σ^2).

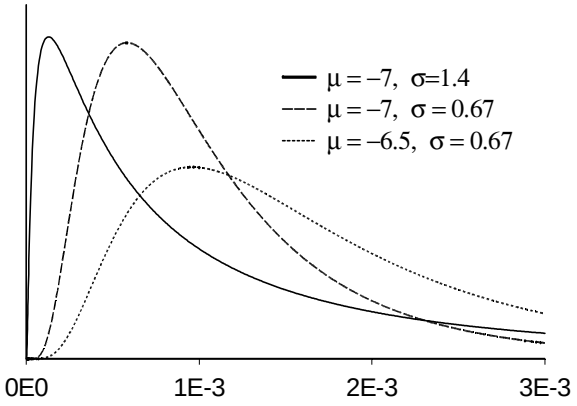


Figure A.10 Three lognormal densities.

A.7.4 The Exponential Distribution

The **exponential distribution** is widely used for modeling time to failure and is inherently associated with the Poisson process [see Martz and Waller (1991)]. For a Poisson random variable X defining the number of failures in a time interval t and for a random variable T defining the time to failure, it can be shown that T has the exponential p.d.f.

$$f(t) = \theta e^{-\theta t},$$

for $t > 0$. Thus, the time to first failure and the times between successive failures follow an exponential distribution and the number of failures in a fixed time interval follows a Poisson distribution.

Figure A.11 shows two exponential densities, for two values of θ. The intercept (height of the curve when $t = 0$) equals θ. Thus, the figure shows that the distribution is more concentrated near zero if θ is large. This agrees with the interpretation of θ as a frequency of failures and t as time to first failure.

The exponential distribution parameter, θ, corresponds to the θt parameterization of the Poisson p.d.f. in Equation A.8 and is referred to as the **failure rate** if the component or system is repaired and restarted immediately after each failure. It is called the **hazard rate** if the component or system can only fail once and cannot be repaired. Section A.4.4.2 discusses modeling duration times with different distributions and defines the hazard rate as $h(t) = f(t)/[1 - F(t)]$. For the exponential distribution, the hazard rate is constant, θ. It can be shown that the exponential distribution is the *only* distribution with a constant hazard rate.

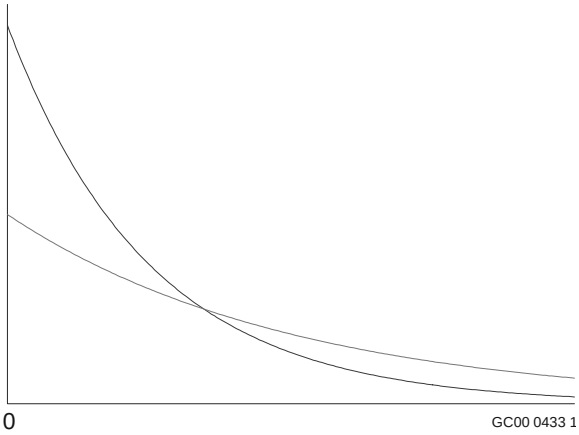


Figure A.11 Two exponential densities.

The c.d.f. of the exponential distribution is

$$F(t) = 1 - e^{-\theta t}.$$

The exponential distribution with parameter θ is denoted exponential(θ). The mean and variance of an exponential(θ) distribution are

$$E(T) = 1/\theta$$

and

$$\mathrm{Var}(T) = 1/\theta^2.$$

The relationship of the exponential distribution to the Poisson process can be seen by observing that the probability of no failures before time t can be viewed in two ways. First, the number of failures, X, can be counted. The probability that the count is equal to 0 is given by Equation A.8 as

$$\Pr(X = 0) = e^{-\lambda t}\frac{(\lambda t)^0}{0!} = e^{-\lambda t}.$$

Alternatively, the probability that first failure time, T, is greater than t is

$$\begin{aligned}\Pr(T > t) &= 1 - \Pr(T \le t) \\ &= 1 - F(t) \\ &= 1 - [1 - e^{-\theta t}] \\ &= e^{-\theta t}.\end{aligned}$$

Thus, the two approaches give the same expression for the probability of no failures before time *t*.

The assumptions of a Poisson process require a constant failure rate, θ, which can be interpreted to mean that the failure process has no memory (Martz and Waller 1991). Thus, if a device is still functioning at time *t*, it remains as good as new and its remaining life has the same exponential(θ) distribution. This constant failure rate corresponds to the flat part of the common **bathtub** curve (frequency of failures plotted against time) and does not pertain to initial (burn-in) failures and wear-out failures.

A different, sometimes useful, parameterization uses $\lambda = 1/\theta = E(T)$. For example, if *T* represents a time to failure, λ is called the mean time to failure. If *T* is the time to repair, or to fire suppression, or to some other event, the name for λ is the mean time to repair, or other appropriate name. The exponential(λ) distribution for *T* has density

$f(t) = (1/\lambda)\exp(-t/\lambda)$, for $t \geq 0$

and c.d.f.

$F(t) = 1 - \exp(-t/\lambda)$, for $t \geq 0$.

The units of λ are the same as the units of *t*, minutes or hours or whatever the data have. The mean and variance are

$E(T) = \lambda$
$\text{var}(T) = \lambda^2$.

A.7.5 The Weibull Distribution

The **Weibull distribution** is widely used in reliability and PRA and generalizes the exponential distribution to include nonconstant failure or hazard rates (Martz and Waller 1991). Different Weibull distributions have been successfully used to describe initial failures and wear-out failures. The Weibull distribution is appropriate when a system is composed of a number of components, and system failure is due to any one of the components failing. It, therefore, is commonly referred to as a distribution corresponding to failure of the weakest link.

For a random variable, *T*, that has a Weibull distribution, the p.d.f. is

$$f(t) = \frac{\beta}{\alpha}\left(\frac{t-\theta}{\alpha}\right)^{\beta-1}\exp\left[-\left(\frac{t-\theta}{\alpha}\right)^{\beta}\right],$$

for $t \geq \gamma \geq 0$ and parameters $\alpha > 0$ and $\beta > 0$. The parameter γ is a location parameter and corresponds to a period of guaranteed life that is not present in many applications (Martz and Waller 1991). Thus, γ is usually set to zero. The c.d.f. for *T* is

$$F(t) = 1 - \exp\left[-\left(\frac{t-\theta}{\alpha}\right)^{\beta}\right],$$

for $t \geq \gamma$ and $\alpha > 0$ and $\beta > 0$.

The α parameter is a scale parameter that expands or contracts the density along the horizontal axis. The β parameter is a shape parameter that allows for a wide variety of distribution shapes. [See Martz and Waller (1991) for further discussion and examples.] When $\beta = 1$, the distribution reduces to the exponential distribution. Therefore, the Weibull family of distributions includes the exponential family of distributions as a special case.

A Weibull distribution with parameters α, β, and γ is referred to as Weibull(α, β, γ) and, when $\gamma = 0$, Weibull(α, β). The mean and variance of the Weibull distribution are given by Martz and Waller (1991) as

$\gamma + \alpha\Gamma(1 + 1/\beta)$

and

$\alpha^2[\Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta)]$.

Here, Γ is the gamma function, defined in Section A.7.6.

Figure A.12 shows four Weibull densities, all with the same scale parameter, α, and all with location parameter $\gamma = 0$. The shape parameter, β, varies. When $\beta < 1$, the density becomes infinite at the origin. When $\beta = 1$, the distribution is identical to the exponential distribution. Surprisingly, the distribution is *not* asymptotically normal as β becomes large, although it is approximately normal when β is near 3.

A.7.6 The Gamma and Chi-Squared Distributions

The **gamma distribution** is an extension of the exponential distribution and is sometimes used as a failure time model (Martz and Waller 1991). It is also often used as a prior distribution in Bayesian estimation (see Appendix B) of the failure rate parameter θ from Poisson(θt) or exponential(θ) data. The **chi-squared distribution** is a re-expression of a special case of the gamma distribution.

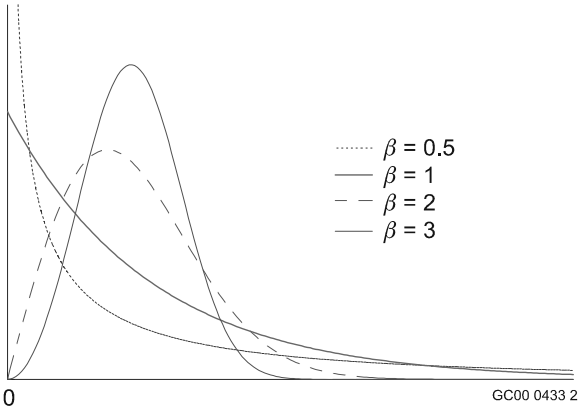


Figure A.12 Four Weibull densities, all having $\gamma = 0$ and all having the same α.

The gamma distribution arises in many ways. The distribution of the sum of independent exponential(θ) random variables is gamma, which forms the basis for a confidence interval for θ from exponential(θ) data. Because the sum of n independent exponentially distributed random variables has a gamma distribution, the gamma distribution is often used as the distribution of the time, or waiting time, to the nth event in a Poisson process. In addition, the chi-squared distribution is the distribution for a sum of squares of independent, identically distributed normal random variables, which forms the basis for a confidence interval for the variance of a normal distribution. The gamma distribution is also often used as a distribution for a positive random variable, similar to the lognormal and Weibull distributions. In PRA work, it is often used as a Bayesian distribution for an uncertain positive parameter.

Two parameterizations of the gamma distribution are common, with various letters used for the parameters. The parameterization given here is most useful for Bayesian updates, the primary use of the gamma distribution in this handbook. For a random variable, T, that has a gamma distribution, the p.d.f. is

$$f(t) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} t^{\alpha-1} \exp(-t\beta),$$

for t, α, and $\beta > 0$.

Here

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

is the **gamma function** evaluated at α. If α is a positive integer, $\Gamma(\alpha) = (\alpha - 1)!$.

A gamma distribution with parameters α and β is referred to as gamma(α, β). The mean and variance of the gamma(α, β) random variable, T, are:

$E(T) = \alpha/\beta$

and

$\mathrm{Var}(T) = \alpha/\beta^2$.

The parameters α and β are referred to as the shape and scale parameters. The shape parameter α allows the density to have many forms. If α is near zero, the distribution is highly skewed. For $\alpha = 1$, the gamma distribution reduces to an exponential(β^{-1}) distribution. Also, the gamma($\alpha = n/2$, $\beta = ½$) distribution is known as the **chi-squared distribution** with n degrees of freedom, denoted $\chi^2(n)$. The p.d.f. for the $\chi^2(n)$ distribution is found by substituting these values into the above formula for the gamma p.d.f. It also can be found in many statistics texts [e.g., Hogg and Craig (1995, Chapter 4)].

In addition, if T has a gamma(α, β) distribution, then $2\beta T$ has a $\chi^2(2\alpha)$ distribution, which forms the defining relationship between the two distributions. The gamma and chi-squared distributions can, therefore, be viewed as two ways of expressing one distribution. Since the chi-squared distribution usually is only allowed to have integer degrees of freedom, the gamma distribution can be thought of as an interpolation of the chi-squared distribution.

Percentiles of the chi-squared distribution are tabulated in Appendix C. These tables can be used as follows to find the percentiles of any gamma distribution. The $100\times p$ percentile of a gamma(α, β) distribution is $\chi^2_p(2\alpha)/(2\beta)$, where $\chi^2_p(2\alpha)$ denotes the $100\times p$ percentile of the chi-squared distribution with 2α degrees of freedom.

Figure A.13 shows gamma densities with four shape parameters, α. When $\alpha < 1$, the density becomes infinite at 0. When $\alpha = 1$, the density is identical to an exponential density. When α is large, the distribution is approximately a normal distribution.

As stated previously, the sum of exponential lifetimes or waiting times has a gamma distribution, with the shape parameter α equal to the number of exponential lifetimes. Also, it has been stated that in general the sum of independent, identically distributed random variables is approximately normal. This is the reason why the gamma distribution is approximately normal when α is large.

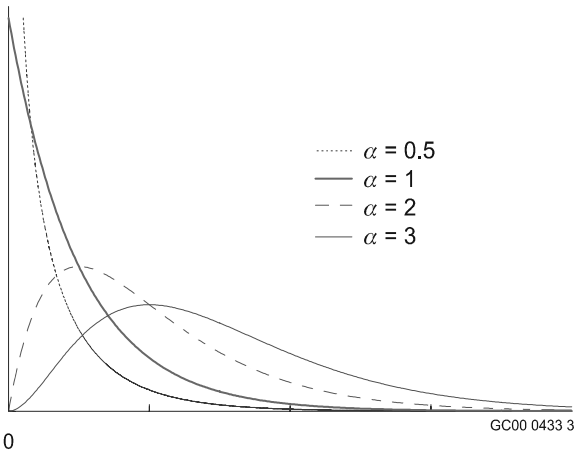


Figure A.13 Gamma densities with four shape parameters.

An alternative parameterization of the gamma distribution uses the scale parameter, say $\lambda = \beta^{-1}$. If T has a gamma(α, λ) distribution, its p.d.f. is

$$f(t) = \frac{1}{\tau^{\alpha}\Gamma(\alpha)} t^{\alpha-1}\exp(-t/\tau)$$

for t, α, and $\lambda > 0$. The mean and variance of the gamma(α, λ) random variable, T, are:

$E(T) = \alpha\lambda$

and

$\mathrm{Var}(T) = \alpha\lambda^2$.

This alternative parameterization is useful in a very small portion of this handbook.

A.7.7 The Inverted Gamma and Inverted Chi-Squared Distributions

The **inverted gamma distribution** is often used as a prior distribution for Bayesian estimation of the mean of an exponential distribution (Martz and Waller 1991). It is also used as a prior and posterior distribution for σ^2 when the data have a normal distribution with variance σ^2 (Box and Tiao 1973, Lee 1997).

For a gamma(α, β) random variable, T, $W = 1/T$ has an inverted gamma distribution with p.d.f.

$$f(w) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\left(\frac{1}{w}\right)^{\alpha+1}\exp\left(-\frac{\beta}{w}\right),$$

for w, α, and $\beta > 0$. The parameters here are the same as for the gamma distribution. For example, if T has units of time then w and β both have units 1/time. A comparison of this density with the gamma density shows that this density has an extra w^2 in the denominator, for reasons explained in Section A.4.7.

The parameters of the inverted gamma distribution are α and β and this distribution is denoted inverted gamma(α, β). Just as with the gamma(α, β) distribution, α is the shape parameter and β is the scale parameter. The distribution can also be parameterized in terms of $\lambda = \beta^{-1}$.

The mean and variance of an inverted gamma(α, β) random variable, W, are

$$E(W) = \frac{\beta}{\alpha - 1}, \qquad \alpha > 1,$$

and

$$\mathrm{Var}(W) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \qquad \alpha > 2.$$

Note that for $\alpha \le 1$ the mean and higher moments do not exist. For $1 < \alpha \le 2$ the mean exists but the variance does not exist (Martz and Waller 1991).

Figure A.14 shows four inverted gamma distributions, all having the same scale parameter, β, and having various shape parameters, α.

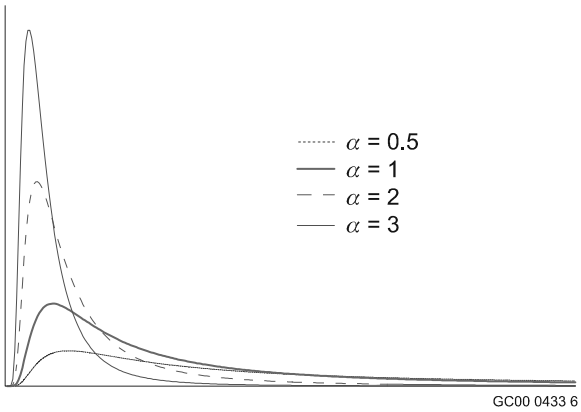


Figure A.14 Four inverted gamma densities, having the same scale parameter, β, and various shape parameters, α.

In the special case with $\alpha = n/2$ and $\beta = ½$, the distribution is called the **inverted chi-squared** distribution with n degrees of freedom. Values from this distribution are sometimes denoted $\chi^{-2}(n)$. This form of the distribution is often used in connection with a prior for σ^2 when the data are normally distributed.

A.7.8 The Beta Distribution

Many continuous quantitative phenomena take on values that are bounded by known numbers a and b. Examples are percentages, proportions, ratios, and distance to failure points on items under stress. The **beta distribution** is a versatile family of distributions that is useful for modeling phenomena that can range from 0 to 1 and, through a transformation, from a to b.

The beta distribution family includes the uniform distribution as well as density shapes that range from decreasing to uni-modal right-skewed to symmetric to U-shaped to uni-modal left-skewed to increasing (Martz and Waller 1991). It can serve as a model for a reliability variable that represents the probability that a system or component lasts at least t units of time. The beta distribution is also widely used in Bayesian estimation and reliability analysis as a prior distribution for the binomial distribution parameter p that represents a reliability or failure probability.

The p.d.f. of a beta random variable, Y, is

$$f(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1},$$

for $0 \# y \# 1$, with the parameters $\alpha, \beta > 0$. The distribution is denoted beta(α, β). The gamma functions at the front of the p.d.f. form a normalizing constant so that the density integrates to 1.

The mean and variance of the beta(α, β) random variable, Y, are

$$E(Y) = \frac{\alpha}{\alpha+\beta}$$

and

$$\mathrm{Var}(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Various beta distributions are shown in Figures A.15 and A.16. Figure A.15 shows beta densities with $\alpha = \beta$, and therefore with mean 0.5. When $\alpha < 1$, the density becomes infinite at 0.0, and when $\beta < 1$, the density becomes infinite at 1.0. When $\alpha = \beta = 1$, the density is uniform. When α and β are large, the density is approximately normal.

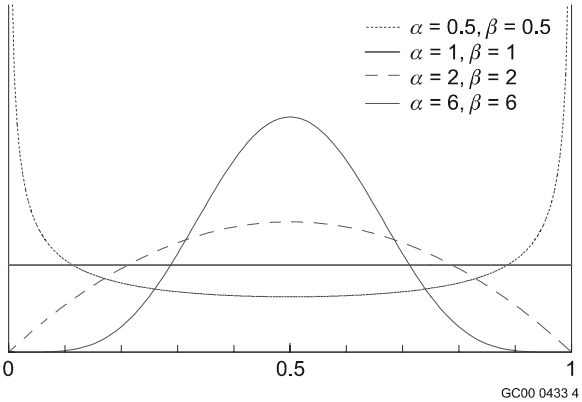


Figure A.15 Beta distributions with mean = 0.5.

Figure A.16 shows densities with mean 0.1. Again, when $\alpha < 1$, the density becomes infinite at 0.0, and when $\alpha > 1$, the density is zero at 0.0. As the parameters α and β become large, the density approaches a normal distribution.

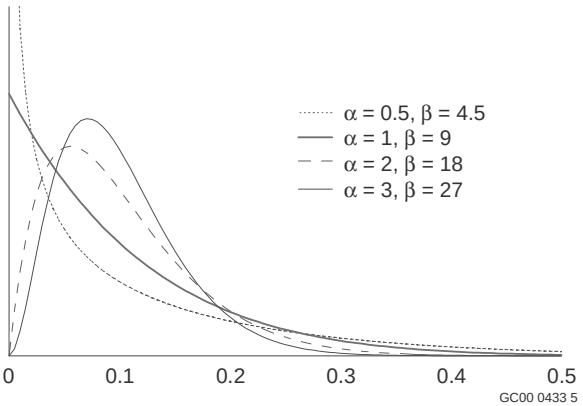


Figure A.16 Four beta distributions with mean 0.1.

Another parameterization of the beta distribution uses the parameters $x_0 = \alpha$ and $n_0 = \alpha + \beta$. This parameterization is used by Martz and Waller (1991) because it simplifies Bayes formulas and Bayesian estimation. The p.d.f. of a beta(x_0, n_0) is

$$f(y) = \frac{\Gamma(n_0)}{\Gamma(x_0)\Gamma(n_0 - x_0)} y^{x_0-1}(1-y)^{n_0-x_0-1},$$

for $0 \# y \# 1$, with the parameters x_0 and n_0 satisfying

$n_0 > x_0 > 0$.

The mean and variance of the beta(x_0, n_0) random variable, Y, are

$$E(Y) = \frac{x_0}{n_0}$$

and

$$\mathrm{Var}(Y) = \frac{x_0(n_0 - x_0)}{n_0^2(n_0 + 1)} .$$

Percentiles of the beta distribution occur in the formula for a confidence interval for p, and in the formula for a Bayes credible interval for p when a conjugate prior is used. Some percentiles are tabulated in Appendix C. In addition, many software packages, including some commonly used spreadsheets, can calculate these percentiles. If none of these work, Martz and Waller (1991) give a method for finding the beta percentiles from the corresponding percentiles of an F distribution, discussed in Section A.7.11. The F distribution is tabulated in most statistics books, and can be interpolated if necessary with good accuracy. The relation is

$$\mathrm{beta}_q(\alpha, \beta) = \alpha / [\alpha + \beta F_{1-q}(2\beta, 2\alpha)]$$

for small q, and

$$\mathrm{beta}_q(\alpha, \beta) = \alpha F_q(2\alpha, 2\beta) / [\beta + \alpha F_q(2\alpha, 2\beta)]$$

for large q. Here $\mathrm{beta}_q(\alpha, \beta)$ denotes the q quantile, or the 100×q percentile, of the beta(α, β) distribution, and $F_q(d_1, d_2)$ denotes the q quantile of an F distribution with d_1 and d_2 degrees of freedom. So if all else fails, and a statistics book with F tables is nearby, the first formula can be used to find the lower percentile of the beta distribution and the second formula can be used to find the upper percentile. This method is not discussed further here, because it is not expected to be needed often.

A.7.9 The Logistic-Normal Distribution

While not widely used in PRA, this distribution is commonly used for Bayesian inference in other fields of application, especially as a prior for the binomial parameter p when p could plausibly be fairly large. X has a logistic-normal distribution if $\ln[X/(1-X)]$ is normally distributed with some mean μ and variance σ^2. The function $\ln[X/(1-X)]$ may appear strange, but it is common enough in some areas of application to have a name, the **logit** function. Therefore, the above statements could be rewritten to say that X has a logistic-normal distribution if logit(X) is normally distributed.

Properties of the logistic-normal distribution are summarized here.

- Let $y = \ln[x/(1-x)]$. Then $x = e^y / (1 + e^y)$. This implies that x must be between 0 and 1.
- As x increases from 0 to 1, $y = \ln[x/(1-x)]$ increases monotonically from $-\infty$ to $+\infty$. Thus, y can be generated from a normal distribution with no problem of forcing x outside its possible range.
- The monotonic relation between x and y means that the percentiles match. For example, the 95th percentile of Y is $\mu + 1.645\sigma$. Denote this by $y_{0.95}$. Therefore, the 95th percentile of X is $x_{0.95} = \exp(y_{0.95}) / [1 + \exp(y_{0.95})]$. Alternatively, this can be written as $y_{0.95} = \ln[x_{0.95} / (1 - x_{0.95})]$.
- If X is close to 0 with high probability, so that $X/(1-X)$ is close to X with high probability, then the logistic-normal and lognormal distributions are nearly the same.

The third bullet shows how to find the percentiles of a logistic-normal distribution. Unfortunately there is no equally easy way to find the moments, such as the mean or variance. Moments must be found using numerical integration.

Figure A.17 shows several logistic normal distributions that all have median 0.5. These correspond to a normally distributed y with mean $\mu = 0$ and with various values of σ. Figure A.18 shows several logistic normal distributions that all have median 0.1. These correspond to a normally distributed y with mean $\mu = -2.2 = \ln[0.1/(1-0.1)]$.

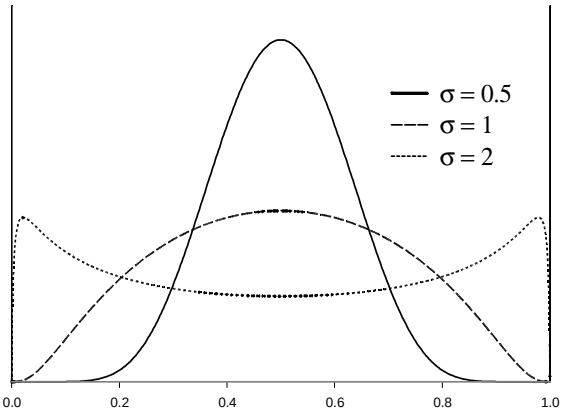


Figure A.17 Three logistic-normal densities with median = 0.5.

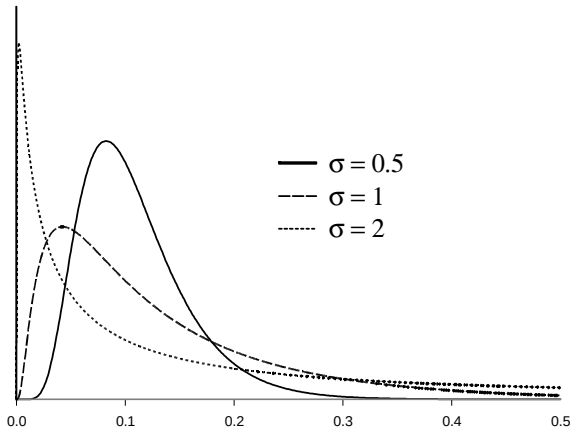


Figure A.18 Three logistic-normal densities with median = 0.1.

Note the general similarities to the beta distributions in Figures A.15 and A.16. Note also the differences: Logistic-normal distributions are characterized most easily by percentiles, whereas beta distributions are characterized most easily by moments. Also, the beta densities can be J-shaped or U-shaped, but the logistic-normal densities always drop to zero at the ends of the range.

A.7.10 Student's *t* Distribution

The **Student's *t*** distribution is not used in a central way in PRA. However, it appears in a peripheral way in places in this handbook, when dealing with the parameters of a normal or lognormal distribution, or in large-sample situations when a distribution is approximated as normal or lognormal. Therefore, the basic facts are summarized here.

If (1) Z has a standard normal distribution, (2) X has a chi-squared distribution with d degrees of freedom, and (3) Z and X are statistically independent, then

$$T = \frac{Z}{\sqrt{X / d}}$$

has a Student's t distribution with d degrees of freedom. Therefore, T has a distribution that is symmetrical about 0, and it can take values in the entire real line. If d is large, the denominator is close to 1 with high probability, and T has approximately a standard normal distribution. If d is smaller, the denominator adds extra variability, and the extreme percentiles of T are farther out than are the corresponding normal percentiles. Tables of the distribution are given in Appendix C.

Although not needed for ordinary work, the p.d.f. and first two moments of T are given here. [See many standard texts, such DeGroot (1975) or Bain and Engelhartd (1992).] The p.d.f. is

$$f(t) = \frac{\Gamma[(d+1)/2]}{(d\pi)^{1/2}\Gamma(d/2)}\left[1+(t^2/2)\right]^{-(d+1)/2} .$$

If $d > 1$ the mean is 0. If $d > 2$ the variance is $d/(d-2)$. If $d \# 2$ the variance does not exist. If $d = 1$, even the mean does not exist; in this case the distribution is called a **Cauchy distribution**.

A.7.11 *F* Distribution

The F distribution, also called Snedecor's F distribution, arises as follows. If Y and Z are independent chi-squared random variables with m and n degrees of freedom, respectively, then

$$X = \frac{Y/m}{Z/n}$$

has an F distribution with m and n degrees of freedom. This is sometimes written as an $F(m, n)$ distribution. This can be re-expressed in terms of a ratio of gamma-distributed variables, because the chi-squared distribution is a special case of a gamma distribution.

The density of an F distribution is almost never needed, although it is given in mathematical statistics books as

$$f(x) = \frac{\Gamma[(m+n)/2]m^{m/2}n^{n/2}}{\Gamma(m/2)\Gamma(n/2)}\frac{x^{(m/2)-1}}{(mx+n)^{(m+n)/2}}$$

for $x \$ 0$. Bain and Engelhardt (1992) give the moments:

$$E(X) = n/(n-2)$$

$$\mathrm{Var}(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} .$$

The mean is defined only if $n > 2$, and the variance only if $n > 4$.

It follows from the definition in terms of a ratio of chi-squared variables that the percentiles are related to each other as follows. If $F_q(m, n)$ is the q quantile (that is, the $100q$ percentile) of an $F(m, n)$ distribution, then

$$F_q(m, n) = 1/F_{1-q}(n, m) . \qquad (A.11)$$

The *F* distribution is also related to the beta distribution, and Equation (A.11) forms the basis for the two different forms of the relation given near the end of Section A.7.8.

The distribution is not tabulated in Appendix C for two reasons: the distribution is used only minimally for the applications in this handbook, and the percentiles and probabilities are given by many commonly used software packages.

A.7.12 Dirichlet Distribution

The Dirichlet distribution is a multivariate generalization of the beta distribution. Let *m* variables $Y_1, ..., Y_m$ be such that $\sum_i Y_i = 1$. Their distribution can be described in terms of any $m - 1$ of them, such as $Y_1, ..., Y_{m-1}$ with

$$Y_m = 1 - \sum_{i=1}^{m-1} Y_i \quad .$$

The *m* variables have a Dirichlet distribution with parameters $\alpha_1, ..., \alpha_m$ if the joint density of the first $m - 1$ variables is

$$f(y_1,...,y_{m-1}) = \frac{\Gamma(\alpha_1+...+\alpha_m)}{\Gamma(\alpha_i)\cdots\Gamma(\alpha_m)} \times y_1^{\alpha_1-1}\cdots y_{m-1}^{\alpha_{m-1}-1}\left(1-\sum_{i=1}^{m-1} y_i\right)^{\alpha_m-1}$$

Observe that when $m = 2$ this reduces to a beta distribution for Y_1 with parameters α_1 and α_2. (Some authors say that $Y_1, ..., Y_m$ have the Dirichlet distribution, while others say that $Y_1, ..., Y_{m-1}$ have this distribution. The distribution is the same whichever way it is described.)

Many of the properties of the distribution are described most easily in terms of an additional parameter α, defined as $\alpha = \alpha_1 + ... + \alpha_m$. Some of these properties are the following.

Individually, each Y_i has a beta(α_i, $\alpha - \alpha_i$) distribution. Therefore, we have

$E(Y_i) = \alpha_i / \alpha$, and

$\text{Var}(Y_i) = \alpha_i(\alpha - \alpha_i)/[\alpha^2(\alpha + 1)]$.

It can also be shown that the covariance terms are given by

$\text{Cov}(Y_i, Y_j) = -\alpha_i \alpha_j/[\alpha^2(\alpha + 1)]$.

Thus, the ratio of each α_i to α determines the corresponding mean. Once the means are fixed, the magnitude of α determines the variances and covariances, with large α corresponding to small variances. The covariances are negative, meaning that if one variable is larger than its mean, each other variable tends to be smaller than its mean; this is not surprising for variables that must sum to 1.

One application of the Dirichlet distribution in PRA is to multiple-branch nodes in event trees. If an event tree has a node with *m* branches, $m > 2$, the probability of the *i*th branch (also called the *i*th "split fraction") can be denoted p_i. The probabilities must satisfy $p_1 + ... + p_m = 1$. They are not known exactly, and therefore are assigned a joint distribution that describes their uncertainty in a Bayesian way. The Dirichlet distribution is a natural distribution to use.

For further information about this distribution, see the article in the *Encyclopedia of Statistical Sciences*, or Kotz, Balakrishnan, and Johnson (2000).

REFERENCES FOR APPENDIX A

Ascher, H., and H. Feingold, 1984, *Repairable Systems Reliability – Modeling, Inference, Misconceptions and Their Causes*, Marcel Dekker, Inc., New York.

Bain, L. J., and M. Engelhardt, 1992, *Introduction to Probability and Mathematical Statistics, 2nd Ed.*, PWS-Kent Pub., Boston.

Box, G. E. P., and G. C. Tiao, 1973, *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading, MA.

Çinlar, E., 1975, *Introduction to Stochastic Processes.* Prentice-Hall, Englewood Cliffs, NJ.

DeGroot, Morris H., 1975, *Probability and Statistics*, Addison-Wesley, Reading, MA.

Derman, C., L. J. Gleser, and I. Olkin, 1973, *A Guide to Probability Theory and Application.* Holt, Rinehart and Winston, NY.

Encyclopedia of Statistical Sciences, Vol. 2, 1982, S. Kotz and N. L. Johnson, eds., John Wiley and Sons, NY, pp. 386-387.

Esary, J. D., and F. Proschan, 1963, Coherent Structures of Non-Identical Components, in *Technometrics*, Vol. 6, pp. 191-209.

Esary, J. D., and F. Proschan, 1970, A Reliability Bound for Systems of Maintained, Interdependent Components in *Journal of the American Statistical Association*, Vol. 65, pp. 329-338.

Hogg, R. V., and A. T. Craig, 1995, *Introduction to Mathematical Statistics, 5th Edition*. Macmillan, NY.

Kotz, S., N. Balakrishnan, and N. L. Johnson, 2000, *Continuous Multivariate Distributions, Volume 1, Models and Applications, 2nd Edition*, Chapter 49. Wiley-Interscience, NY.

Lee, Peter M., 1997, *Bayesian Statistics: An Introduction, Second Edition*, Arnold, a member of the Hodder Headline Group, London.

Martz, H. F., and R. A. Waller, 1991, *Bayesian Reliability Analysis*. R. E. Krieger Publishing Co., Malabar, FL.

Meyer, P. L., 1970, *Introductory Probability and Statistical Applications*. Addison-Wesley Pub. Co., Reading, MA.

NRC, 1994, *A Review of NRC Staff Uses of Probabilistic Risk Assessment*, NUREG-1489, U. S. Nuclear Regulatory Commission, Washington, D.C.

Pfeiffer, P. E., and D. A. Schum, 1973, *Introduction to Applied Probability*. Academic Press, NY.

Snedecor, G. W., and W. G. Cochran, 1989, *Statistical Methods*. Iowa State University Press, Ames, IA.

Thompson, W. A., Jr., 1981, On the Foundations of Reliability, in *Technometrics*, Vol. 23, No. 1, pp. 1-13.

Vesely, W. E., F. F. Goldberg, N. H. Roberts, and D. F. Haasl, 1981, *Fault Tree Handbook*. NUREG-0492. U.S. Nuclear Regulatory Commission, Washington, D.C.

B. BASICS OF STATISTICS

B.1 Random Samples

When sampling from a distribution (or population), it is usually assumed that the n observations are taken at random, in the following sense. It is assumed that the n random variables $X_1, X_2, ..., X_n$ are independent. That is, the sample $X_1, X_2, ..., X_n$, taken from a distribution $f(x)$, has the joint p.d.f. h satisfying

$$h(x_1, x_2, ..., x_n) = f(x_1) \cdot f(x_2) \cdot \ldots \cdot f(x_n).$$

This follows the definition of independent random variables given in Section A.4.5. A sample taken in this way is called a **random sample**. (As elsewhere in this handbook, upper case letters denote random variables and lower case letters denote particular values, number.)

The random variables $X_1, X_2, ..., X_n$ forming such a random sample are referred to as being independent and identically distributed. If n is large enough, the sampled values will represent the distribution well enough to permit inference about the true distribution.

B.2 Sample Moments

Mathematical expectation and moments provide characteristics of distributions of random variables. These ideas can also be used with observations from a random sample from a distribution to provide **estimates** of the parameters that characterize that distribution.

A **statistic** is a function of one or more random variables that does not depend on any unknown parameters. A function of random variables that can be computed from the collected data sample is thus a statistic. Note that a function of random variables is also a random variable that has its own probability distribution and associated characteristics.

If $X_1, X_2, ..., X_n$ denote a random sample of size n from a distribution $f(x)$, the statistic

$$\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

is the **mean of the random sample**, or the **sample mean** and the statistic

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1} \qquad \text{(B.1)}$$

is the **variance of the random sample**. Note that n ! 1 is used as the denominator in the S^2 statistic to make the statistic an *unbiased* estimator of the population variance, σ^2 (unbiased estimators are discussed in Section B.4.1). Some authors use n in the denominator instead of n ! 1, with corresponding adjustment of formulas that involve S, but this handbook uses Equation B.1 consistently. In applications with computer packages, note which definition is used and make any necessary adjustments to formulas in this handbook.

Although not used as much as the sample mean and sample variance, the **sample skewness** is occasionally of interest. The definition can vary in detail, but one, used by SAS (1988) is

$$\frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} (X_i - \overline{X})^3 \Big/ S^3 \ .$$

Similarly, the statistics defined by

$$m_r = \sum_{i=1}^{n} \frac{X_i^r}{n},$$

for r = 1, 2, ..., are called the **sample moments**.

One of the common uses of statistics is estimating the unknown parameters of the distribution from which the sample was generated. The sample mean, or average, $\overline{X}$, is used to estimate the distribution mean, or population mean, $\therefore$, the sample variance, S^2, is used to estimate the population variance, σ^2, and so forth.

B.3 Statistical Inference

Since values of the parameters of a distribution are rarely known, the distribution of a random variable is rarely completely known. However, with some assumptions and information based on a random sample of observations from the distribution or population, values of the unknown parameters can often be estimated. Probabilities can then be calculated from the corresponding distribution using these parameter estimates.

Statistical inference is the area of statistics concerned with using sample data to answer questions and make statements about the distribution of a random variable from which the sample data were obtained. Parameter **estimators** are functions of sample data that are used to estimate the distribution parameters. Statements about parameter values are inferred from the specific sample to the general distribution of the random variable or population. This inference cannot be perfect; all inference techniques involve uncertainty. Understanding the performance properties of various estimators has received much attention in the statistics field.

For the purposes of this handbook, statistical inference procedures can be classified as follows:

- parameter estimation
  - estimation by a point value
  - estimation by an interval
- hypothesis testing
  - tests concerning parameter values
  - goodness-of-fit tests and other model-validation tests.

Parametric statistical inference assumes that the sample data come from a particular, specified family of distributions, with only the parameter values unknown. However, not all statistical inference is based on parametric families. In many cases, in addition to not knowing the distribution parameter values, the form of the parametric family of distributions is unknown. **Distribution-free**, also called **nonparametric**, techniques are applicable no matter what form the distribution may have. **Goodness-of-fit tests** are an important type of nonparametric tests that can be used to test whether a data set follows a hypothesized distribution.

For statistical inference, two major approaches exist, the **frequentist** approach and the **Bayesian** approach. The two resulting sets of inference tools are summarized in Sections B.4 and B.5. In PRA work, Bayesian estimators are normally used for parameter estimation. See, for example, NUREG-1489 (NRC 1994). However, frequentist hypothesis tests are often used for model validation, especially when the hypothesis to be tested does not involve a simple parameter. This use of Bayesian techniques for estimation and frequentist techniques for model validation is also recommended by Box (1980).

NUREG-1489 (NRC 1994) lists a number of "advantages" and "disadvantages" for each of the Bayesian and frequentist approaches. An "advantage" is often in the eye of the beholder. For example, is it an advantage or disadvantage that frequentist methods use only the data at hand, not external or prior information? Therefore, the lists from that report are presented in modified and augmented form in Table B.1, where the points are not called advantages or disadvantages, but simply "features."

B.4 Frequentist Inference

Frequentist estimation of distribution parameters uses only the information contained in the data sample and assumptions about a model for the sample data. In contrast to **Bayesian estimation** (discussed in Section B.5), degree of belief is not incorporated into the estimation process of frequentist estimation.

In the **frequentist** approach to estimation, a distribution parameter is treated as an unknown constant and the data to be used for estimation are assumed to have resulted from a random sample. Information outside that contained in the sample data is used minimally. The random variability in the sample data is assumed to be due directly to the process under study. Thus, the frequentist approach addresses variation in parameter estimates and how far estimates are from the true parameter values.

Frequentist testing of a hypothesis follows the same spirit. The hypothesis is assumed, and the data are compared to what would have been expected or predicted by the hypothesis. The frequentist analyst asks whether the observed values come from the likely part of the distribution or from the extreme tails, and decides in this way whether the data are consistent with the hypothesis.

B.4.1 Point Estimation

Many situations arise in statistics where a random variable X has a p.d.f. that is of known functional form but depends on an unknown parameter θ that can take on any value in a set. The different values for θ produce a family of distributions. One member of the family corresponds to each possible value of θ. **Estimators** of the distribution parameter are functions of sample data that are used to estimate the distribution parameters. Thus, estimators are themselves random variables. The specific value of the estimator computed from a random sample provides an **estimate** of the distribution parameter. Note the distinction between *estimator*, a random variable, and *estimate*, a particular value. An estimate of a distribution parameter in the form of a single number is called a **point estimate** of that parameter. The sample mean is a point estimate of

Table B.1 Features of Bayesian and frequentist approaches.

Bayesian Approach	Frequentist Approach
Bayesian methods allow the formal introduction of prior information and knowledge into the analysis, which can be especially useful when sample data are scarce, such as for rare events. For the nuclear industry, this knowledge often exists in the form of **industry-wide generic data**. Thus, Bayesian estimation allows the use of various types of relevant generic data in PRA.	Results depend only on the data sample. Including relevant information about a parameter that is external to the random sample is complicated.
If the prior distribution accurately reflects the uncertainty about a parameter, Bayesian parameter estimates are better than classical estimates.	
Bayesian estimation can be sensitive to the choice of a prior distribution. Therefore: Identifying suitable prior distributions and justifying and gaining acceptance for their use can be difficult. The choice of a prior distribution is open to criticism that the choice is self-serving and may reflect inappropriate, biased, or incorrect views.	
Because Bayesian probability intervals can be interpreted as probability statements about a parameter, they are easily combined with other sources of uncertainty in a PRA using the laws of probability.	A confidence interval cannot be directly interpreted as a probability that the parameter lies in the interval.
Bayesian distributions can be propagated through fault trees, event trees, and other logic models.	It is difficult or impossible to propagate frequentist confidence intervals through fault and event tree models common in PRA to produce corresponding interval estimates on output quantities of interest.
Using Bayes' Theorem, Bayesian estimation provides a method to update the state of knowledge about a parameter as additional data become available.	Frequentist methods can update an earlier analysis if the original data are still available or can be reconstructed.
In complicated settings, Bayesian methods require software to produce samples from the distributions.	In complicated settings, frequentist methods must use approximations. In some cases they may be unable to analyze the data at all.
Bayesian hypothesis tests are commonly used only with hypotheses about a parameter value.	A well-developed body of hypothesis tests exists, useful for model validation. These are appropriate for investigating goodness of fit, poolability of data sources, and similar questions that do not involve a simple parameter.
Both Approaches	
When the quantity of data is large, both approaches produce good estimates.	
Both types of computation are straightforward when estimating a parameter in a simple setting.	

the mean of the distribution and the sample variance is a point estimate of the variance of the distribution. For another sample drawn from the same population, a different sample mean and variance would be calculated. In fact, these sample statistics are specific values of random variables. Thus, viewed as random variables the sample statistics have their own **sampling distributions**. For example, it can be shown that $\overline{X}$ has mean μ and variance σ^2/n, regardless of the distribution from which the samples are drawn.

Different techniques exist for obtaining point estimates for unknown distribution characteristics or parameters. Two of the most common methods are presented here [see Hogg and Craig (1995) for more information]: maximum likelihood estimation and the method of moments.

A distribution of a random variable X that depends on an unknown parameter θ will be denoted $f(x; \theta)$. If X_1, X_2, ..., X_n is a random sample from $f(x; \theta)$, the joint p.d.f. of X_1, X_2, ..., X_n is $f(x_1; \theta) \cdot f(x_2; \theta) \cdot \ldots \cdot f(x_n; \theta)$. This joint p.d.f. may be viewed as a function of the unknown parameter θ and, when so viewed, is called the **likelihood function**, L, of the random sample. Thus, the likelihood function is the joint p.d.f. of X_1, X_2, ..., X_n, denoted

$$L(\theta; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \theta) ,$$

viewed as a function of θ. The **maximum likelihood estimate** of θ is defined as the value $\hat{\theta}$ such that $L(\hat{\theta}; x_1, x_2, ..., x_n) \geq L(\theta; x_1, x_2, ..., x_n)$ for every value of θ. That is, the maximum likelihood estimate of θ is the value $\hat{\theta}$ that maximizes the likelihood function. In many cases, this maximum will be unique and can often be obtained through differentiation. Note that solving the derivative set to zero for θ may be easier using $\ln(L)$, which is equivalent since a function and its natural logarithm are maximized at the same value of θ.

The maximum likelihood estimate is a function of the observed random sample x_1, x_2, ..., x_n. When $\hat{\theta}$ is considered to be a function of the random sample X_1, X_2, ..., X_n, then $\hat{\theta}$ is a random variable and is called the **maximum likelihood estimator** of θ.

Another method of point estimation is the **method of moments**, which involves setting the distribution moments equal to the sample moments:

$$M_r = E(X^r) = m_r = \sum x_i^r / n,$$

for $r = 1, 2, ..., k$, if the p.d.f. $f(x; \theta_1, \theta_2, ..., \theta_k)$ has k parameters. The k equations can be solved for the k unknowns $\theta_1, \theta_2, ..., \theta_k$ and the solutions $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ are the method-of-moments estimators.

How "well" a point estimator estimates a parameter has received a large amount of attention. Numerous desirable properties of point estimators exist. One desirable property of estimators, alluded to previously in Section B.2, is **unbiasedness**. An **unbiased** estimator is one whose mean value is equal to the parameter being estimated. That is, an estimator $\hat{\theta}$ is unbiased for a parameter θ if $E(\hat{\theta}) = \theta$. For a random sample from a normal distribution, the sample mean, $\overline{X}$, and the sample variance, S^2, are unbiased estimators of μ and σ^2, respectively. (Recall that S^2 is defined by Equation B.1, with $n - 1$ in the denominator.) However, the method of moments estimator of the variance is biased. The **bias** of an estimator $\hat{\theta}$ is defined as $E(\hat{\theta}) - \theta$.

Minimum variance is another desirable property of an estimator. An unbiased estimator is said to have minimum variance if its variance is less than or equal to the variance of every other unbiased statistic for θ. Such an estimator is referred to as an unbiased, minimum variance estimator.

Another desirable property of estimators is **sufficiency**. For a random sample X_1, X_2, ..., X_n from $f(x; \theta_1, \theta_2, ..., \theta_m)$, and $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$ functions (statistics) of the X_is, the statistics $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$ are **jointly sufficient statistics** if the conditional p.d.f. of the X_is given the statistics $\hat{\theta}_j$s, $g(x_1, x_2, ..., x_n | \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m)$, is independent of the parameters (Martz and Waller, 1991).

Sufficiency can be thought of as exhausting all the possible information about a parameter that is contained in the random sample. When a sufficient statistic exists, it may serve as the basis for a minimum variance or "best" estimator of the parameter. Sufficiency is also important because it simplifies Bayesian estimation methods.

Under certain commonly occurring conditions, as the sample size gets large, the maximum likelihood estimator is approximately normally distributed, approximately unbiased, and has approximately the minimum variance. It is, therefore, a very good estimator for large data sets. The maximum likelihood estimator is not necessarily good for small data sets.

Several other methods of estimation and desirable properties for estimators exist. Further information can be found in Hogg and Craig (1995) or Kendall and Stuart (1973).

B.4.2 Interval Estimation

Another way of estimating a parameter is to identify that it falls in some interval (*lcl*, *ucl*) with a specified degree of certainty, or confidence, where *lcl* denotes the lower confidence limit and *ucl* denotes the upper confidence limit. The interval (*lcl*, *ucl*) is referred to as an **interval estimate** of the parameter. The *lcl* and *ucl* values are calculated from the random sample from the given distribution. Associating a level of desired confidence with an interval estimate produces a **confidence interval**. The level of desired confidence is also referred to as the confidence coefficient.

Confidence intervals are based on estimators associated with a random sample (functions of the data), *LCL* for the lower confidence limit and *UCL* for the upper confidence limit, such that, prior to observing the random sample, the probability that the unknown parameter, θ, is contained in the interval [*LCL*, *UCL*] is known. That is,

$\Pr[LCL \leq \theta \leq UCL] = 1 - \alpha$

for $0 < \alpha < 1$.

Once the random sample has been generated, the functions *LCL* and *UCL* produce two values, *lcl* and *ucl*. The interval (*lcl*, *ucl*) is called a two-sided confidence interval with confidence level $1 - \alpha$, or equivalently, a $100(1 - \alpha)\%$ two-sided confidence interval. Similarly, upper one-sided confidence intervals or lower one-sided confidence intervals can be defined that produce only an upper or lower limit, respectively.

Since the true parameter value, although unknown, is some constant, the interval estimate either contains the true parameter value or it does not. A 95% confidence interval is interpreted to mean that, for a large number of random samples from the same distribution, 95% of the resulting intervals (one interval estimate of the same population parameter constructed the same way for each sample) would contain the true population parameter value, and 5% of the intervals would not. The $\alpha = .05$ risk of obtaining an interval that does not contain the parameter can be increased or decreased. Values for $1 - \alpha$ should be decided upon prior to obtaining the random sample, with .99, .95, and .90 being typical. Note that higher confidence levels result in wider interval estimates.

Confidence intervals cannot be interpreted as probability statements about the parameter being estimated, because the parameter is assumed to be an unknown constant and not a random variable. The level of confidence pertains to the percentage of intervals, each calculated from a different random sample from the same distribution, that are expected to contain the true parameter value. The confidence does not pertain to the specific calculated interval (it could be from the unlucky 5% of intervals that do not contain the true parameter value).

As an example, a confidence interval for the parameter μ can be produced from a random sample drawn from a normal(μ, σ^2) population by calculating the appropriate functions of the data. Recall that, if each sample value is drawn from a normal distribution, the sample mean $\bar{X}$ has a normal(μ, σ^2/n) distribution, where n is the sample size. Even if the sample values are drawn from a distribution that is not normal, by the central limit theorem, $\bar{X}$ will be approximately normal(μ, σ^2/n) for sufficiently large n. Assuming that σ^2 is known (from previous data and experience), the standardized normal random variable

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

is normal(0, 1), and tabulated in Appendix C. From these tables, values of w can be found for which

$$\Pr[-w \leq Z \leq w] = 1 - \alpha. \qquad \text{(B.2)}$$

For example, for $\alpha = .05$, $w = 1.96$. In this case, w is the 97.5th percentile of the standard normal distribution, commonly denoted $z_{0.975}$, or $z_{1-\alpha/2}$ for $\alpha = .05$.

Substituting for Z in Equation B.2 above, along with some algebraic manipulation, produces

$$\Pr[\bar{X} - w\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + w\frac{\sigma}{\sqrt{n}}] = 1 - \alpha,$$

which defines a $100(1 - \alpha)\%$ confidence interval for the population mean μ, where

$$LCL = \bar{X} - w\frac{\sigma}{\sqrt{n}} \qquad \text{(B.3)}$$

and

$$UCL = \bar{X} + w\frac{\sigma}{\sqrt{n}}, \qquad \text{(B.4)}$$

with $w = z_{1-\alpha/2}$.

A random sample will yield a specific sample mean. The numbers w and n are known, and σ was assumed to be known. Therefore, for a preassigned confidence level, values for *LCL* and *UCL* can be calculated to produce a specific 100(1 − α)% confidence interval for μ. Each of the random variables *LCL* and *UCL* is a statistic, and the interval (*LCL*, *UCL*) is a random interval formed from these statistics.

Usually the value of σ is not known. In this case, the unbiased estimator of the population variance, S^2, can be used to produce S, which can be used in the above equations in place of σ. Thus, the following standardized random variable, T, can be formed:

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

(This formula requires the definition of S based on Equation B.1.) For sufficiently large n (say 25 or 30), T follows a normal(0, 1) distribution. If n is not sufficiently large, T follows a **Student's *t* distribution**, for which tabulated probabilities exist in many statistics books, and in Appendix C. The Student's t distribution depends on a parameter called the **degrees of freedom**. In the present example, this parameter equals $n - 1$. Confidence intervals for the population mean can then be calculated similarly to the case where σ is known, using either the Student's t distribution or, when n is large, the normal distribution.

Confidence intervals can also be constructed for differences of means and many other population parameters, such as variances, probabilities, quantiles, and distribution characteristics (see, for example, Hogg and Craig 1978).

B.4.3 Hypothesis Testing

Testing a statistical hypothesis is another major area of statistics. A **hypothesis** is a statement about the distribution of the observable random variable. Often this statement is expressed as a statement about one or more parameters of the distribution. As discussed previously, estimation uses information in the data from a random sample to infer something about the magnitude of a parameter value. Similar to estimation, **hypothesis testing** also uses information from the random sample. However, the objective of hypothesis testing is to determine whether the specific statement about the distribution is true.

The hypothesis to be tested is referred to as the **null hypothesis**, denoted by H_0. The alternative to the null hypothesis is referred to as the **alternative hypothesis**, denoted H_1 or H_a. A **test of a hypothesis** is a rule or procedure for deciding whether to reject or accept the null hypothesis. This rule or procedure is based upon information contained in the random sample and produces a single number, called a test statistic, which leads to a decision of whether the sample values do not support H_0. The entire set of values that the test statistic may assume is divided into two regions, one corresponding to the **rejection region** and the other to the **acceptance region**.

If the test statistic computed from a particular sample has a value in the rejection region, H_0 is rejected. If the test statistic falls in the acceptance region, H_0 is said to be accepted, due to lack of evidence to reject. For each of the two possible cases for H_0, true or false, the test either rejects or does not reject H_0, producing four distinct possibilities. These possibilities (using conditional probability notation), along with some concepts and terms associated with hypothesis testing, are summarized in Table B.2 (Martz and Waller, 1991).

Table B.2 Possible hypothesis test outcomes.

	H_0 True	H_0 False
Accept H_0	Pr(accept H_0 | H_0 is true) = 1 − α = Level of confidence	Pr(accept H_0 | H_0 is false) = β = Pr(Type II Error)
Reject H_0	Pr(reject H_0 | H_0 is true) = α = Level of significance = Pr(Type I Error)	Pr(reject H_0 | H_0 is false) = 1 − β = Power

A stated null hypothesis is either true or false. One of two errors can occur in hypothesis testing:

1. rejection of the null hypothesis when it is true, referred to as the **Type I error**; and
2. acceptance of the null hypothesis when it is false, referred to as the **Type II error**.

The probability of making a Type I error, denoted by α, is referred to as the **significance level of the test**. Thus, $1 - \alpha$ is the probability of making a correct decision when H_0 is true. The probability of making a correct decision when H_0 is false, denoted $1 - \beta$, is referred to as the **power of the test**. The probability of making a Type II error is equal to one minus the power of the test, or β.

The goodness of a statistical hypothesis test is measured by the probabilities of making a Type I or a Type II error. Since α is the probability that the test statistic will fall in the rejection region, assuming H_0 to be true, increasing the size of the rejection region will increase α and simultaneously decrease β for a fixed sample size. Reducing the size of the rejection region will decrease α and increase β. If the sample size, n, is increased, more information will be available for use in making the decision, and both α and β will decrease.

The probability of making a Type II error, β, varies depending on the true value of the population parameter. If the true population parameter is very close to the hypothesized value, a very large sample would be needed to detect such a difference. That is, the probability of accepting H_0 when H_0 is false, β, varies depending on the difference between the true value and the hypothesized value. For hypothesis tests, α is specified prior to conducting the random sample. This fixed α specifies the rejection region. For a deviation from the hypothesized value that is considered practical and that is wished to be detectable by the hypothesis test, a sample size can be selected that will produce an acceptable value of β.

Different alternative hypotheses will result in different rejection regions for the same H_0. This is seen most easily for a hypothesis that is expressed in terms of a parameter, for example, H_0: $\theta = \theta_0$ for some given value θ_0. In this case, there is an exact correspondence between one-sided and two-sided confidence intervals and rejection regions for one-sided and two-sided alternative hypotheses. If the hypothesized value falls outside a $100(1 - \alpha)\%$ confidence interval for the corresponding population parameter, the null hypothesis would be rejected with level of confidence equal to $1 - \alpha$.

For the example presented in the previous section, Section B.4.2, the $100(1 - \alpha)\%$ two-sided confidence interval for a population mean is defined by the *LCL* and *UCL* in Equations B.3 and B.4. For the hypothesized value of the mean, say θ_0, if $\theta_0 < lcl$ or $\theta_0 > ucl$, H_0 would be rejected. Equivalently, the test statistic in Equation B.2 can be computed using $\theta = \theta_0$ and, for $\alpha = .05$, if it is greater than 1.96 or less than -1.96, H_0 would be rejected with 95% level of confidence.

To further illustrate these concepts, a more detailed example is presented. Atwood et al. (1998) assert that for non-momentary losses of offsite power with plant-centered causes, the recovery times are lognormally distributed with median 29.6 minutes and error factor 10.6. This is equivalent to X being normally distributed with $\theta = \ln(29.6) = 3.388$ and $\sigma = \ln(10.6)/1.645 = 1.435$, where $X = \ln$(recovery time in minutes). Suppose that a plant of interest has experienced five such losses of offsite power in recent history. It is desired to test whether the plant's recovery times follow the claimed distribution.

To simplify the situation, the question is formulated in terms of θ only, assuming that $\sigma = 1.435$. The null hypothesis is

H_0: $\theta = 3.388$.

Because only long recovery times are of concern from a risk standpoint, the alternative hypothesis is defined as

H_1: $\theta > 3.388$.

That is, values <3.388 are possible, but are not of concern. The test statistic, based on $n = 5$ recovery times, is to reject H_0 if

$$Z = \frac{\overline{X} - 3.388}{1.435/\sqrt{5}} > w \quad .$$

To make α, the probability of Type I error, equal to 0.05, w is chosen to be the 95th percentile of the standard normal distribution, 1.645. Then the test can be re-expressed as rejecting H_0 if

$\overline{X} > 4.44$.

The upper part of Figure B.1 shows the density of $\overline{X}$ when $\theta = 3.388$. The area to the right of 4.44 is

$\Pr(\overline{X} > 4.44 \mid H_0 \text{ is true})$,

which equals 0.05.

What if H_0 is false? For example, a median 60-minute recovery time corresponds to $\mu = \ln(60) = 4.09$. The lower part of Figure B.1 shows the density of $\bar{X}$ when $\mu = 4.09$. The area to the right of 4.44 is

$$\Pr(\bar{X} > 4.44 \mid \mu = 4.09),$$

which is equal to 0.29. This value represents the power of the hypothesis test when $\mu = 4.09$ and is the probability of (correctly) rejecting H_0. The probability of a Type II error when $\mu = 4.09$ is 1 ! 0.29 = 0.71.

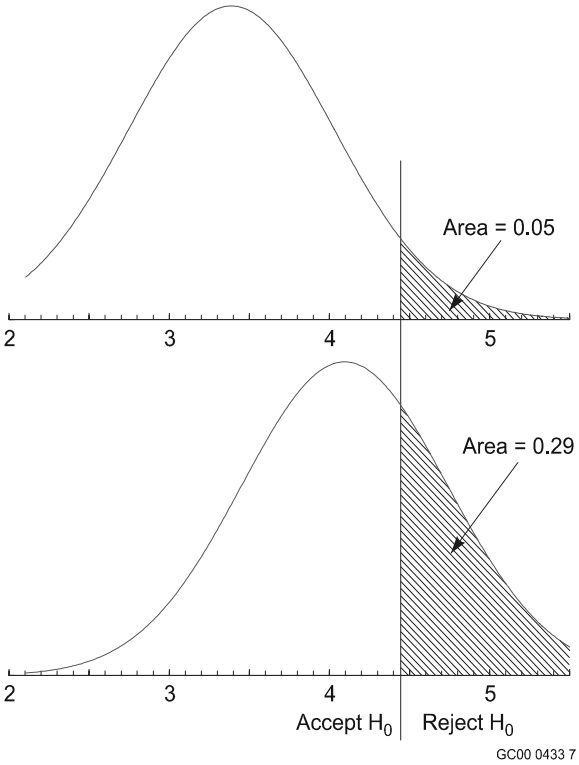


Figure B.1 Probability of rejecting H_0: $\mu = 3.388$, if in fact H_0 is true (upper distribution), and if H_0 is false with $\mu = 4.09$ (lower distribution).

It can be useful to plot the power as a function of μ. The plot is called a **power curve**. Figure B.2 shows two power curves, corresponding to $n = 5$ and $n = 10$. The probability of Type I error, that is, the probability of rejecting H_0 when H_0 is true, is shown as α. The probability of Type II error, that is, the probability of accepting H_0 when H_0 is false, is shown as β for one value of μ, and equals 1 minus the power. The two tests, with $n = 5$ and $n = 10$, have both been calibrated so that $\alpha = 0.05$. The power, for any value of μ in H_1, is larger when $n = 10$ than when $n = 5$; equivalently, the probability of Type II error is smaller.

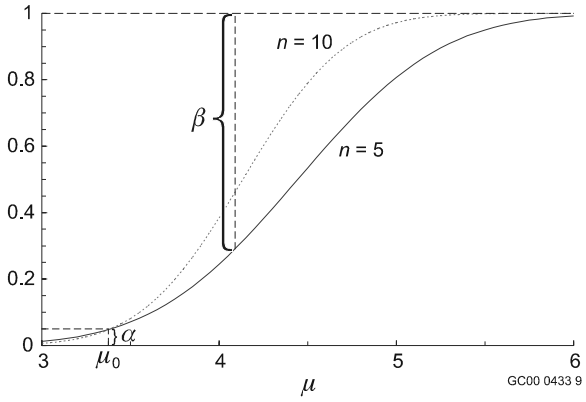


Figure B.2 Power curves when $n = 5$ and $n = 10$. The graph shows the probability of rejecting H_0, as a function of the true μ.

The interpretation of confidence in hypothesis testing is also the same as with confidence intervals. That is, the confidence is not in one specific test statistic. The confidence arises from the viewpoint that if the random sample was collected a number of times in the same way and if H_0 was true, 100(1 ! α)% of the tests would result in not rejecting H_0.

As can be seen, interval estimation and hypothesis testing are closely related. Some experimenters prefer expressing inference as estimators. Others prefer to test a particular hypothesized value for the parameter of interest.

B.4.4 Goodness-of-Fit Tests

The methods presented above are concerned with estimating the parameters of a distribution, with the actual form of the distribution assumed to be known (or the central limit theorem applies with large n). Other hypothesis tests do not assume that only a parameter is unknown. In particular, **goodness-of-fit tests** are special hypothesis tests that can be used to check on the assumed distribution itself. Based on a random sample from some distribution, goodness-of-fit tests test the hypothesis that the data are distributed according to a specific distribution. In general, these tests are based on a comparison of how well the sample data agree with an expected set of data from the assumed distribution.

Perhaps the most familiar goodness-of-fit test is the **chi-square** test. The test statistic used for this statistical test has an approximate χ^2 distribution, leading to the name of the test. A random sample of n observations, $X_1, X_2, \ldots, X_n$, can be divided or binned into k groups or intervals, referred to as bins, producing an empirical

distribution. The assumed distribution under the null hypothesis, $f_0(x)$, is used to calculate the probability that an observation would fall in each bin, with the probabilities denoted by $p_1, p_2, ..., p_k$.

These probabilities are frequently referred to as cell probabilities. The k bins are also called cells. The k bin intervals do not overlap and they completely cover the range of values of $f_0(x)$. It follows that $\sum_{i=1}^{k} p_i = 1$. The *expected* frequency of the ith bin, denoted e_i, is $e_i = np_i$, for $i = 1, 2, ..., k$. The e_i are commonly referred to as the expected cell counts. The *observed* frequencies for each of the k bins, denoted O_i, are referred to as observed cell counts.

The chi-square goodness-of-fit test compares the observed frequencies to the corresponding expected frequencies for each of the k groups by calculating the test statistic:

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} .$$

If the observations come from some distribution other than that specified in the null hypothesis, the observed frequencies tend to agree poorly with the expected frequencies, and the computed test statistic, X^2, becomes large.

The distribution of the quantity X^2 can be approximated by a chi-square distribution. The parameter that specifies the chi-square distribution is called the **degrees of freedom**. Its value depends on the number of unknown parameters and how they are estimated. When the null-hypothesis distribution is completely specified, such as normal with both μ and σ known, the degrees of freedom are $k - 1$. If, instead, H_0 specifies the form of the distribution but not the parameters, the degrees of freedom must be adjusted. In the example, if $\bar{X}$ and S^2 from the sample are used to estimate μ and σ^2 when testing the distribution, the degrees of freedom are between $k - 1$ and $k - 1 - m$, where m is the number of estimated parameters, 2. If the quantity X^2 is greater that the $1 - \alpha$ quantile of the $\chi^2(k - 1)$ distribution, the hypothesized probability distribution is rejected. If X^2 is less than the $1 - \alpha$ quantile of the $\chi^2(k - 1 - m)$ distribution, the data are concluded to be adequately modeled by $f_0(x)$.

When the sample size is small, the χ^2 distribution still applies as long as the expected frequencies are not too small. Larger expected cell counts make the chi-square distribution approximation better. The problem with small expected frequencies is that a single random observation falling in a group with a small expected frequency would result in that single value having a major contribution to the value of the test statistic, and thus, the test itself. In addition, small expected frequencies are likely to occur only in extreme cases. One rule of thumb is that no expected frequency should be less than 1 (see Snedecor and Cochran, 1989). Two expected frequencies can be near 1 if most of the other expected frequencies are greater than 5. Groups with expected frequencies below 1 should be combined or the groups should be redefined to comply with this rule. Note that k is the number of groups after such combination or redefinition.

Comparing how well sample data agree with an expected set of data leads to another common use of the chi-square test: testing whether two or more classification criteria, used to group subjects or objects, are independent of one another. Although not a goodness-of-fit test, the **chi-square test for independence** is similar to the chi-square goodness-of-fit test.

For two grouping criteria, the rows of a two-way **contingency table** can represent the classes of one of the criteria and the columns can represent the classes of the other criterion. To test the hypothesis that the rows and columns represent independent classifications, the expected number, e_{ij}, that would fall into each cell of the two-way table is calculated and used to compute the following chi-square test statistic:

$$X^2 = \sum_{i,\,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} ,$$

where $i = 1, 2, ..., r$ (the number of rows); $j = 1, 2, ..., c$ (the number of columns); and O_{ij} is the number observed to belong to the ith row and jth column. The e_{ij} are calculated by

$$e_{ij} = \frac{R_i C_j}{n} ,$$

where R_i and C_j are the total observed in the ith row and jth column, respectively, and n is the total sample size ($n = \sum R_i = \sum C_j$).

For this test, the χ^2 test statistic follows a chi-square distribution with $(r - 1)(c - 1)$ degrees of freedom. If the calculated X^2 exceeds the $1 - \alpha$ quantile of the χ^2 distribution with $(r - 1)(c - 1)$ degrees of freedom, the null hypothesis of independence is rejected and the rows and columns are concluded to not represent independent classifications.

The **Kolmogorov goodness-of-fit test** tests the hypothesis that the observed random variable has c.d.f. $F_0(x)$, versus the alternative hypothesis that the observed random variable does not have c.d.f. $F_0(x)$. It does this by comparing the sample c.d.f. (the empirical distribution function) to the hypothesized c.d.f. For a random sample of n observations, $X_1, X_2, ..., X_n$, the test statistic is defined as the maximum vertical distance between the empirical c.d.f., $\hat{F}(x)$ and $F_0(x)$. The actual procedure for calculating the test statistic can be found in many statistics texts, including Martz and Waller (1991) and Conover (1999). The test statistic is then compared to the 1 ! " quantile of tabled values for the Kolmogorov test, e.g. in Table C. If the calculated test statistic exceeds the 1 ! " quantile, the hypothesized c.d.f. is rejected. Otherwise, $F_0(x)$ is concluded to describe the data. The Kolmogorov goodness-of-fit test is based on each individual data point and therefore is equally effective for small or large samples.

As an example, consider the previous example of loss-of-offsite-power recovery times. Suppose that five recovery times have been observed at the plant: 7, 22, 94, 185, and 220 minutes. The corresponding values of x = ln(recovery time in minutes) are 1.95, 3.09, 4.54, 5.22, and 5.39. The null hypothesis and alternative hypothesis are:

H_0: X is normal with μ = 3.388, σ = 1.435
H_1: X has some other distribution .

Note, all possible alternative distributions are considered, not just normal distributions, or distributions with σ = 1.435.

Figure B.3 shows the distribution function specified by H_0 (the smooth curve) and the empirical distribution function specified by the data (the step function). The maximum distance between the two distributions is D, the Kolmogorov test statistic. If D is large, the test rejects H_0 in favor of H_1.

If the sample size is small, the Kolmogorov test may be preferred over the chi-square test. The Kolmogorov test is exact, even for small samples, while the chi-square test is an approximation that is better for larger sample sizes. The Kolmogorov statistic can also be used to construct a confidence region for the unknown distribution function.

The Kolmogorov goodness-of-fit test is sometimes called the Kolmogorov-Smirnov one-sample test. Statistics that are functions of the maximum vertical distance between $\hat{F}(x)$ and $F_0(x)$ are considered to be Kolmogorov-type statistics. Statistics that are functions of the maximum vertical distance between *two* empirical distribution functions are considered to be Smirnov-type statistics. A test of whether two samples have the same distribution function is the Smirnov test, which is a two-sample version of the Kolmogorov test presented above. This two-sample test is also called the **Kolmogorov-Smirnov two-sample test**. Conover (1999) presents additional information and tests.

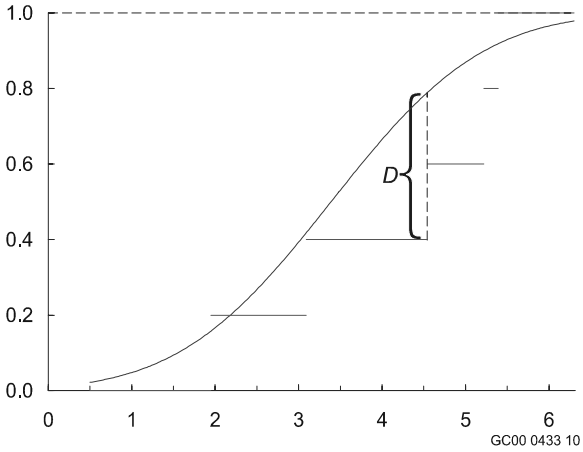


Figure B.3 The hypothesized distribution, the empirical distribution, and the Kolmogorov test statistic, D.

Another useful goodness-of-fit test is the **Anderson-Darling** goodness-of-fit test and test for normality. The Anderson-Darling test measures the squared difference between the empirical distribution function (EDF) of a sample and the theoretical distribution to be tested. It averages this squared difference over the entire range of the random variable, weighting the tails more heavily than the center. This statistic is recommended to guard against wayward observations in the tail and has generally good power.

Because many statistical methods require the assumption of normality, some assessment of whether data come from a normal population is helpful when considering appropriate analysis techniques. The Anderson-Darling statistic provides a measure of how much normal probability scores for the data (normal probability plot values) deviate from a straight line that would arise under normality. A computer package is often used to calculate this statistic and compare it to tabled values for the statistic. If the calculated statistic is too high, the deviations from the straight line are too large to be attributed to the variation due to sampling observations from a normal distribution. Thus, the hypothesis of normality is rejected. See Stephens (1982) for more information on the Anderson-Darling goodness-of-fit test.

Certain patterns of deviations from linearity in normal probability plots indicate common types of nonnormal characteristics, such as skewness or kurtosis (presence of long or short tails of the p.d.f.). Test for skewness or kurtosis are also available. See Snedecor and Cochran (1989) for more information on these tests.

B.5 Bayesian Estimation

B.5.1 Purpose and Use

Bayesian estimation is the other major class of statistical inference methods. Similar to frequentist estimation, both point and interval estimates can be obtained. However, Bayesian estimation is different from classical estimation in both practical and philosophical perspectives (NRC, 1994). Bayesian estimation incorporates degree of belief and information beyond that contained in the data sample, forming the practical difference from classical estimation. The subjective interpretation of probability forms the philosophical difference from frequentist methods.

The prior belief about a parameter's value is contained in what is referred to as the **prior distribution**, which describes the state of knowledge (or subjective probability) about the parameter, prior to obtaining the data sample. Therefore, in the Bayesian approach, the parameters of the sampling distribution have probability distributions. These probabilities do not model random variability of the parameters, but the analyst's degree of belief about the true values of the unknown parameters. The distributions are sometimes called "uncertainty distributions." A Bayesian uncertainty distribution satisfies all the rules of a probability distribution.

Bayesian estimation consists of two main areas, both of which use the notion of subjective probability. The first area involves using available data to assign a subjective, prior distribution to a parameter, such as a failure rate. The degree of belief about the uncertainty in a parameter value is expressed in the prior distribution. This assignment of a prior distribution does not involve the use of Bayes' Theorem. The second area of Bayesian estimation involves using additional or new data to update an existing prior distribution. This is called **Bayesian updating**, and directly uses Bayes' Theorem.

Bayes' Theorem, presented in Section A.5, can be seen to transform the prior distribution by the effect of the sample data distribution to produce a **posterior distribution**. The sample data distribution, $f(x^*2)$, can be viewed as a function of the unknown parameter, instead of the observed data, x_i, producing a likelihood function, as discussed in Section B.4.1. Using the likelihood function, the fundamental relationship expressed by Bayes' Theorem is

$$\text{Posterior Distribution} = \frac{\text{Prior Distribution} \times \text{Likelihood}}{\text{Marginal Distribution}}.$$

The marginal distribution serves as a normalizing constant.

In Bayesian updating, the sampling distribution of the data provides new information about the parameter value. Bayes' Theorem provides a mathematical framework for processing new sample data as they become sequentially available over time. With the new information, the uncertainty of the parameter value has been reduced, but not eliminated. Bayes' Theorem is used to combine the prior and sampling distributions to form the posterior distribution, which then describes the updated state of knowledge (still in terms of subjective probability) about the parameter. **Point and interval estimates** of the parameter can then be obtained directly from the posterior distribution, which is viewed as containing the current knowledge about the parameter. This posterior distribution can then be used as the prior distribution when the next set of data becomes available. Thus, Bayesian updating is successively implemented using additional data in conjunction with Bayes' Theorem to obtain successively better posterior distributions that model plant-specific parameters.

Bayesian point and interval estimates are obtained from both the prior and posterior distributions. The interval estimates are **subjective probability intervals**, or **credible intervals**. The terminology is not yet universally standard. Berger (1985) and Bernardo and Smith (2000) both use the term **credible interval**, but Box and Tiao (1973) use **Bayes probability interval**, Lindley (1965) uses **Bayesian confidence interval**, and other authors have used other terms. A credible interval can be interpreted as a subjective probability statement about the parameter value, unlike classical interval estimates. That is, the interpretation of a 95% Bayesian posterior probability interval (a, b) is that, with 95% subjective probability, the parameter is contained in the interval (a, b), given the prior and sampling distributions.

B.5.2 Point and Interval Estimates

Bayesian parameter estimation involves four steps. The first step is identification of the parameter(s) to be estimated, which involves consideration of the assumed distribution of the data that will be collected. The second step is development of a prior distribution that

appropriately quantifies the state of knowledge concerning the unknown parameter(s). The third step is collection of the data sample. The fourth and final step is combining the prior distribution with the data sample using Bayes' Theorem, to produce the desired posterior distribution.

For PRA applications, determining the prior distribution is usually based on generic data and the data sample usually involves site-specific or plant-specific operating data. The resulting posterior distribution would then be the site-specific or plant-specific distribution of the parameter.

The plant-specific data collected are assumed to be a random sample from an assumed sampling distribution. The data are used to update the prior, producing the posterior distribution. Point estimates, such as the most likely value (the mode), the median, or (most commonly) the mean value, and probability interval estimates of the parameter can then be obtained. Other bounds and other point values can also be obtained with the Bayesian approach because the posterior parameter distribution is entirely known and represents the available knowledge about the parameter.

Bayesian interval estimation is more direct than classical interval estimation and is based solely on the posterior p.d.f.. A symmetric 100(1 ! ")% **two-sided Bayes probability interval estimate** of the parameter is easily obtained from the "/2 and 1 ! "/2 quantiles of the posterior distribution. **Lower and upper one-sided Bayes probability interval estimates** can similarly be calculated. Again, note that the Bayes interval estimates are explicit probability statements of the true parameter being contained in the interval.

In some applications, such as a planned facility, plant-specific data do not exist. In these cases, Bayes' Theorem is not used. Only the generic data are used and parameter estimates are based solely on the assumed prior distribution. Investigation of the sensitivity of the results to the choice of the prior distribution is important for these cases.

B.5.3 Prior Distributions

The prior distribution is fundamental to any Bayesian analysis and represents all that is known or assumed about the parameter 2 prior to collecting any data. The information summarized by the prior distribution can be objective, subjective, or both. Operational data and data from a previous but comparable experiment could be used as objective data. Subjective information could involve personal experience and opinions, expert judgement, assessments of degree of belief, and design information.

The selection of prior distributions can be seen to be somewhat subjective. A particular prior must be evaluated to determine the sensitivity of the choice of that prior on the parameter estimates. Consistency of the prior information and data with the prior distribution must be tested.

Choices for the initial prior distribution and techniques for handling various kinds of data are described in detail in several references, such as Martz and Waller (1991), Raiffa and Schlaifer (1961), and Siu and Kelly (1998) .

B.5.3.1 Noninformative Prior Distributions

One class of prior distributions that is widely used is termed **noninformative priors**, also referred to as priors of ignorance, or **reference priors** (Bernardo and Smith 1994). These names refer to the situation where very little *a priori* information about a parameter is available in comparison to the information expected to be provided by the data sample, or there is indifference about the range of values the parameter could assume.

One might think that this indifference could be expressed by a prior distribution that is uniformly distributed over the interval of interest. Every value in the interval is equally likely and no knowledge about any specific value over another value is imposed.

However, uniform distributions do not necessarily best reflect true noninformativeness (Box and Tiao 1973), because models can be parameterized in various ways. For example, if the time to failure, *T*, is exponentially distributed, it is common to write the density of *T* as

$$f(t) = \lambda e^{-\lambda t}$$

or alternatively as

$$f(t)=\frac{1}{\mu}e^{-t/\mu}.$$

The two parameters are related by 8 = 1/ ∴.

A uniform distribution cannot be said to automatically reflect ignorance and be used as a standard noninformative prior distribution. For the exponential example here, ignorance of 8 implies ignorance of ∴, but 8 and ∴ cannot both have a uniform distribution. In fact, suppose that 8 has the uniform distribution in

some finite range, say from a to b. Then λ has a density proportional to $1/\lambda^2$ in the range from $1/b$ to $1/a$, as stated in Appendix A.4.7. The distribution of λ is *not* uniform.

Jeffreys' rule (Jeffreys 1961) guides the choice of noninformative prior distributions and provides the Jeffreys prior distribution (Box and Tiao, 1973). The Jeffreys prior distribution is commonly used in PRA and involves using a specific parameterization of the model (distribution). Jeffreys' method is to transform the model into a parameterization that is in terms of a **location parameter**, a parameter that slides the distribution sideways without changing its shape. (See Box and Tiao 1978, Secs. 1.2.3 and 1.3.4). This method then uses the uniform distribution as the noninformative prior for the location parameter. It is reasonable to regard a uniform distribution as noninformative for a location parameter. The distribution for any other parameterization is then determined, and is called noninformative.

In the exponential example, working with log(time), let $\theta = \log(\lambda)$, $S = \log(T)$, and $s = \log(t)$. Using algebraic formulas given in Section A.4.7 of Appendix A, it can be shown that the density in this parameterization is

$$f(s) = \exp(s-\theta)e^{-\exp(s-\theta)} .$$

Because θ only appears in the expression $s - \theta$, a change in θ simply slides the distribution sideways along the s axis. Therefore, θ is a location parameter. The Jeffreys noninformative prior is a uniform distribution for θ. This distribution translates to a density for β which is proportional to $1/\beta$, and a density for λ which is proportional to $1/\lambda$. These are the Jeffreys noninformative prior distributions for β and λ.

A further argument for Jeffreys prior distributions is that the resulting Bayes intervals are numerically equal to confidence intervals (Lindley 1958), and the confidence intervals are based on the data alone, not on prior belief. Unfortunately, the above approach cannot be followed exactly when the data come from a discrete distribution, such as binomial or Poisson. The original parameter can only approximately be converted to a location parameter. The resulting distribution is still called the Jeffreys prior, however, even though it only approximates the Jeffreys method.

To avoid the appearance of pulling prior distributions out of the air, the general formula for the Jeffreys prior is stated here, as explained by Box and Tiao (1973) and many others. All the particular cases given in this handbook can be found by working out the formula in those cases. Let θ denote the unknown parameter to be estimated. Let $L(\theta; x)$ denote the likelihood corresponding to a single observation. It is a function of θ, but it also depends on the data, x. For example, x is the number of Poisson events in a single unit of time, or the number of failures on a single demand, or the length of a single duration. Calculate

$$-\frac{d^2}{d\theta^2}\ln[L(\theta;x)] .$$

Now replace the number x by the random variable X, and evaluate the expectation of the calculated derivative:

$$E\left(-\frac{d^2}{d\theta^2}\ln[L(\theta;X)]\right) .$$

The Jeffreys noninformative prior is a function of θ proportional to the square root of this expectation.

B.5.3.2 Conjugate Prior Distributions

A **conjugate prior distribution** is a distribution that results in a posterior distribution that is a member of the same family of distributions as the prior. Therefore, conjugate prior distributions are computationally convenient. The methodology for obtaining conjugate priors is based on sufficient statistics and likelihood functions (see Martz and Waller, 1991).

The beta family of distributions is the conjugate family of prior distributions for the probability of failure of a component in a binomial sampling situation. The resulting posterior beta distribution can then be used to provide point and interval estimates of the failure probability.

A time-to-failure random variable is often assumed to follow an exponential distribution, with the failure events arising from a Poisson process. For this model, with either exponential or Poisson data, the gamma family of distributions is the conjugate family of prior distributions for use in Bayesian reliability and failure rate analyses.

Figure B.4 is a schematic diagram showing the relation of the kinds of priors that have been mentioned so far. There are many nonconjugate priors. A relatively small family of priors is conjugate, typically a single type such as the gamma distributions or beta distributions. Finally, the Jeffreys noninformative prior is a single

distribution, shown in the diagram by a dot. In many cases, the Jeffreys prior is also conjugate, as indicated in the figure.

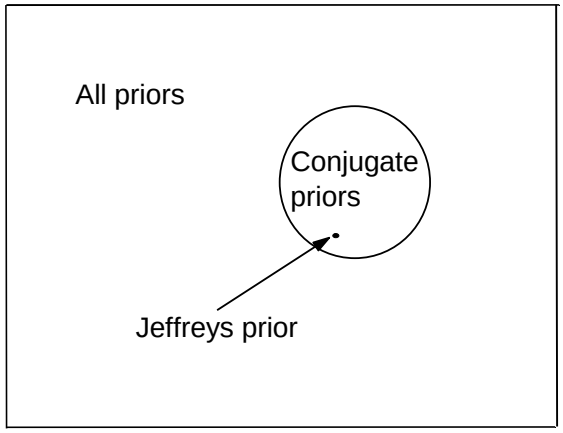


Figure B.4 Schematic diagram of types of priors.

A popular nonconjugate prior is the lognormal distribution. It can be used as a prior distribution for both the binomial sampling and Poisson process models above, although it is not conjugate.

Conjugate prior distributions provide convenience, but accurate modeling of prior degree of belief should not be sacrificed for mathematical convenience. However, when one expression of prior belief is viewed to be as correct as another, the more convenient expression is usually selected for use.

B.5.3.3 Other Prior Distribution Approaches

The prior distribution is the distribution of degree of belief before data that provide new information are obtained. Usually, the prior probabilities do not have a direct frequency interpretation and cannot be experimentally confirmed.

When the prior distribution does have a frequency interpretation, the observed data can be used to estimate the prior distribution. This situation represents another class of methods of statistical inference called **empirical Bayes** methods. The empirical Bayes prior distribution is empirically determined, for example, using observed plant-specific data for a given set of plants. Bayes' Theorem can then be applied to combine this prior with observed data from a specific plant to produce a posterior distribution. Thus, empirical Bayes methods are useful when data from similar, but not identical, sources exist. This situation also gives rise to the use of so-called **hierarchical Bayes** methods (see Gelman, et al., 1995, and Carlin and Louis, 1996).

Attempts have been made to remove some of the subjectivity present in selecting prior distributions, with the goal being to obtain *one* distribution for the same given information. That is, different analysts using the same information would decide upon the same prior distribution. The result has been development of the **method of maximum entropy**. If θ is a parameter with uncertainty distribution g, the entropy is defined as

$$-\int g(\theta)\ln[g(\theta)]d\theta \quad .$$

The distribution g that maximizes this expression is called the maximum entropy distribution. For finite ranges, the p.d.f. with the largest entropy is the uniform, or flat, distribution. Thus, entropy can be viewed as a measure of the variability in the height of a p.d.f., and a maximum entropy prior would be the one with the required mean that is as flat as possible. Siu and Kelly (1998, Table 2) give the maximum entropy distributions for a number of possible constraints.

Maximum entropy methods may see more use in the future, but still do not produce a systematic approach to selecting only one prior from a set of possible priors. In fact, the same problem that the Jeffreys' method attempts to address (Section B.5.3.1) is present with the maximum entropy approach: if a model is parameterized in two different ways, the maximum entropy priors for the two parameters are inconsistent with each other.

To address this lack of invariance, **constrained noninformative priors** are obtained. They are based on the maximum entropy approach in conjunction with Jeffreys' method. That parameterization is used for which the parameter is a location parameter. Giving maximum entropy to this parameter produces a distribution called the constrained noninformative prior distribution. Atwood (1996) presents constrained noninformative priors and their application to PRA. Constrained noninformative prior distributions are seeing use in PRA, although not as much as Jeffreys' priors.

Other ways of defining noninformative prior distributions exist. See Martz and Waller (1991) and Berger (1985) for more information.

REFERENCES FOR APPENDIX B

Atwood, C. L., 1996, Constrained Noninformative Priors in Risk Assessment, in *Reliability Engineering and System Safety*, vol. 53, pp. 37-46.

Atwood, C. L., D. L. Kelly, F. M. Marshall, D. A. Prawdzik and J. W. Stetkar, 1998, *Evaluation of Loss of Offsite Power Events at Nuclear Power Plants: 1980-1996*, NUREG/CR-5496, INEEL/EXT-97-00887, Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.

Berger, J. O., 1985, *Statistical Decision Theory and Bayesian Analysis, Second Edition*, Springer-Verlag, New York.

Bernardo, J. M., and A. F. M. Smith, 1994, *Bayesian Theory*, John Wiley & Sons, New York.

Box, G. E. P., 1980, Sampling and Bayes' Inference in Scientific Modelling and Robustness (with discussion) in *Journal of the Royal Statistical Society, Series A*, Vol. 143, Part 4, pp. 383-430.

Box, G. E. P., and G. C. Tiao, 1973, *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA.

Carlin, B. P., and T. A. Louis, 1996, *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall, London.

Conover, W. J., 1999, *Practical Nonparametric Statistics*, 3rd ed. John Wiley & Sons, NY.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 1995, *Bayesian Data Analysis*, Chapman & Hall, London.

Hogg, R. V., and A. T. Craig, 1995, *Introduction to Mathematical Statistics*, Macmillan, NY.

Jeffreys, H., 1961, *Theory of Probability, Third Ed.*, Clarendon Press, Oxford, England.

Kendall, M. G., and A. Stuart, 1973, *The Advanced Theory of Statistics*, Vol. 2, Hafner, NY.

Lindley, D. V., 1958, Fiducial Distributions and Bayes' Theorem, in *Journal of the Royal Statistical Society, Series B*, Vol. 20, pp. 102-107.

Lindley, D. V., 1965, *Introduction to Probability and Statistics from a Bayesian Viewpoint*, Cambridge University Press, Cambridge, UK.

Martz, H. F., and R. A. Waller, 1991, *Bayesian Reliability Analysis*, R. E. Krieger Publishing Co., Malabar, FL.

NRC, 1994, *A Review of NRC Staff uses of Probabilistic Risk Assessment*, NUREG-1489, U. S. Nuclear Regulatory Commission, Washington, D.C.

Raiffa, H., and R Schlaifer, 1961, *Applied Statistical Decision Theory*, The M.I.T. Press, Cambridge, MA.

SAS, 1988, *SAS Procedures Guide, Release 6.03 Edition*, SAS Institute, Inc., Cary, NC.

Siu, N., and D. Kelly, 1998, Bayesian Parameter Estimation in Probabilistic Risk Assessment, in *Reliability Engineering and System Safety*, Vol. 62, pp. 89-116.

Snedecor, G. W., and W. G. Cochran, 1989, *Statistical Methods*, Iowa State University Press, Ames, IA.

Stephens, M. A., 1982, Anderson-Darling Test for Goodness of Fit, in *Encyclopedia of Statistical Sciences*, Vol. 1., pp. 81-85, S. Kotz and N. L. Johnson, eds., John Wiley & Sons, NY.

C. STATISTICAL TABLES

Table C.1. Standard normal cumulative distribution function, Φ.

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
-4.00	3.2E-5	-2.52	0.0059	-2.02	0.0217	-1.52	0.0643
-3.50	2.3E-4	-2.51	0.0060	-2.01	0.0222	-1.51	0.0655
-3.00	0.0013	-2.50	0.0062	-2.00	0.0228	-1.50	0.0668
-2.99	0.0014	-2.49	0.0064	-1.99	0.0233	-1.49	0.0681
-2.98	0.0014	-2.48	0.0066	-1.98	0.0239	-1.48	0.0694
-2.97	0.0015	-2.47	0.0068	-1.97	0.0244	-1.47	0.0708
-2.96	0.0015	-2.46	0.0069	-1.96	0.0250	-1.46	0.0721
-2.95	0.0016	-2.45	0.0071	-1.95	0.0256	-1.45	0.0735
-2.94	0.0016	-2.44	0.0073	-1.94	0.0262	-1.44	0.0749
-2.93	0.0017	-2.43	0.0075	-1.93	0.0268	-1.43	0.0764
-2.92	0.0018	-2.42	0.0078	-1.92	0.0274	-1.42	0.0778
-2.91	0.0018	-2.41	0.0080	-1.91	0.0281	-1.41	0.0793
-2.90	0.0019	-2.40	0.0082	-1.90	0.0287	-1.40	0.0808
-2.89	0.0019	-2.39	0.0084	-1.89	0.0294	-1.39	0.0823
-2.88	0.0020	-2.38	0.0087	-1.88	0.0301	-1.38	0.0838
-2.87	0.0021	-2.37	0.0089	-1.87	0.0307	-1.37	0.0853
-2.86	0.0021	-2.36	0.0091	-1.86	0.0314	-1.36	0.0869
-2.85	0.0022	-2.35	0.0094	-1.85	0.0322	-1.35	0.0885
-2.84	0.0023	-2.34	0.0096	-1.84	0.0329	-1.34	0.0901
-2.83	0.0023	-2.33	0.0099	-1.83	0.0336	-1.33	0.0918
-2.82	0.0024	-2.32	0.0102	-1.82	0.0344	-1.32	0.0934
-2.81	0.0025	-2.31	0.0104	-1.81	0.0351	-1.31	0.0951
-2.80	0.0026	-2.30	0.0107	-1.80	0.0359	-1.30	0.0968
-2.79	0.0026	-2.29	0.0110	-1.79	0.0367	-1.29	0.0985
-2.78	0.0027	-2.28	0.0113	-1.78	0.0375	-1.28	0.1003
-2.77	0.0028	-2.27	0.0116	-1.77	0.0384	-1.27	0.1020
-2.76	0.0029	-2.26	0.0119	-1.76	0.0392	-1.26	0.1038
-2.75	0.0030	-2.25	0.0122	-1.75	0.0401	-1.25	0.1056
-2.74	0.0031	-2.24	0.0125	-1.74	0.0409	-1.24	0.1075
-2.73	0.0032	-2.23	0.0129	-1.73	0.0418	-1.23	0.1093
-2.72	0.0033	-2.22	0.0132	-1.72	0.0427	-1.22	0.1112
-2.71	0.0034	-2.21	0.0136	-1.71	0.0436	-1.21	0.1131
-2.70	0.0035	-2.20	0.0139	-1.70	0.0446	-1.20	0.1151
-2.69	0.0036	-2.19	0.0143	-1.69	0.0455	-1.19	0.1170
-2.68	0.0037	-2.18	0.0146	-1.68	0.0465	-1.18	0.1190
-2.67	0.0038	-2.17	0.0150	-1.67	0.0475	-1.17	0.1210
-2.66	0.0039	-2.16	0.0154	-1.66	0.0485	-1.16	0.1230
-2.65	0.0040	-2.15	0.0158	-1.65	0.0495	-1.15	0.1251
-2.64	0.0041	-2.14	0.0162	-1.64	0.0505	-1.14	0.1271
-2.63	0.0043	-2.13	0.0166	-1.63	0.0516	-1.13	0.1292
-2.62	0.0044	-2.12	0.0170	-1.62	0.0526	-1.12	0.1314
-2.61	0.0045	-2.11	0.0174	-1.61	0.0537	-1.11	0.1335
-2.60	0.0047	-2.10	0.0179	-1.60	0.0548	-1.10	0.1357
-2.59	0.0048	-2.09	0.0183	-1.59	0.0559	-1.09	0.1379
-2.58	0.0049	-2.08	0.0188	-1.58	0.0571	-1.08	0.1401
-2.57	0.0051	-2.07	0.0192	-1.57	0.0582	-1.07	0.1423
-2.56	0.0052	-2.06	0.0197	-1.56	0.0594	-1.06	0.1446
-2.55	0.0054	-2.05	0.0202	-1.55	0.0606	-1.05	0.1469
-2.54	0.0055	-2.04	0.0207	-1.54	0.0618	-1.04	0.1492
-2.53	0.0057	-2.03	0.0212	-1.53	0.0630	-1.03	0.1515

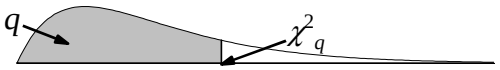
Table C.1 Standard normal cumulative distribution function, M (continued).

z	M(*z*)	*z*	M(*z*)	*z*	M(*z*)	*z*	M(*z*)
-1.02	0.1539	-0.51	0.3050	0.00	0.5000	0.51	0.6950
-1.01	0.1562	-0.50	0.3085	0.01	0.5040	0.52	0.6985
-1.00	0.1587	-0.49	0.3121	0.02	0.5080	0.53	0.7019
-0.99	0.1611	-0.48	0.3156	0.03	0.5120	0.54	0.7054
-0.98	0.1635	-0.47	0.3192	0.04	0.5160	0.55	0.7088
-0.97	0.1660	-0.46	0.3228	0.05	0.5199	0.56	0.7123
-0.96	0.1685	-0.45	0.3264	0.06	0.5239	0.57	0.7157
-0.95	0.1711	-0.44	0.3300	0.07	0.5279	0.58	0.7190
-0.94	0.1736	-0.43	0.3336	0.08	0.5319	0.59	0.7224
-0.93	0.1762	-0.42	0.3372	0.09	0.5359	0.60	0.7257
-0.92	0.1788	-0.41	0.3409	0.10	0.5398	0.61	0.7291
-0.91	0.1814	-0.40	0.3446	0.11	0.5438	0.62	0.7324
-0.90	0.1841	-0.39	0.3483	0.12	0.5478	0.63	0.7357
-0.89	0.1867	-0.38	0.3520	0.13	0.5517	0.64	0.7389
-0.88	0.1894	-0.37	0.3557	0.14	0.5557	0.65	0.7422
-0.87	0.1922	-0.36	0.3594	0.15	0.5596	0.66	0.7454
-0.86	0.1949	-0.35	0.3632	0.16	0.5636	0.67	0.7486
-0.85	0.1977	-0.34	0.3669	0.17	0.5675	0.68	0.7517
-0.84	0.2005	-0.33	0.3707	0.18	0.5714	0.69	0.7549
-0.83	0.2033	-0.32	0.3745	0.19	0.5753	0.70	0.7580
-0.82	0.2061	-0.31	0.3783	0.20	0.5793	0.71	0.7611
-0.81	0.2090	-0.30	0.3821	0.21	0.5832	0.72	0.7642
-0.80	0.2119	-0.29	0.3859	0.22	0.5871	0.73	0.7673
-0.79	0.2148	-0.28	0.3897	0.23	0.5910	0.74	0.7703
-0.78	0.2177	-0.27	0.3936	0.24	0.5948	0.75	0.7734
-0.77	0.2206	-0.26	0.3974	0.25	0.5987	0.76	0.7764
-0.76	0.2236	-0.25	0.4013	0.26	0.6026	0.77	0.7794
-0.75	0.2266	-0.24	0.4052	0.27	0.6064	0.78	0.7823
-0.74	0.2296	-0.23	0.4090	0.28	0.6103	0.79	0.7852
-0.73	0.2327	-0.22	0.4129	0.29	0.6141	0.80	0.7881
-0.72	0.2358	-0.21	0.4168	0.30	0.6179	0.81	0.7910
-0.71	0.2389	-0.20	0.4207	0.31	0.6217	0.82	0.7939
-0.70	0.2420	-0.19	0.4247	0.32	0.6255	0.83	0.7967
-0.69	0.2451	-0.18	0.4286	0.33	0.6293	0.84	0.7995
-0.68	0.2483	-0.17	0.4325	0.34	0.6331	0.85	0.8023
-0.67	0.2514	-0.16	0.4364	0.35	0.6368	0.86	0.8051
-0.66	0.2546	-0.15	0.4404	0.36	0.6406	0.87	0.8078
-0.65	0.2578	-0.14	0.4443	0.37	0.6443	0.88	0.8106
-0.64	0.2611	-0.13	0.4483	0.38	0.6480	0.89	0.8133
-0.63	0.2643	-0.12	0.4522	0.39	0.6517	0.90	0.8159
-0.62	0.2676	-0.11	0.4562	0.40	0.6554	0.91	0.8186
-0.61	0.2709	-0.10	0.4602	0.41	0.6591	0.92	0.8212
-0.60	0.2743	-0.09	0.4641	0.42	0.6628	0.93	0.8238
-0.59	0.2776	-0.08	0.4681	0.43	0.6664	0.94	0.8264
-0.58	0.2810	-0.07	0.4721	0.44	0.6700	0.95	0.8289
-0.57	0.2843	-0.06	0.4761	0.45	0.6736	0.96	0.8315
-0.56	0.2877	-0.05	0.4801	0.46	0.6772	0.97	0.8340
-0.55	0.2912	-0.04	0.4840	0.47	0.6808	0.98	0.8365
-0.54	0.2946	-0.03	0.4880	0.48	0.6844	0.99	0.8389
-0.53	0.2981	-0.02	0.4920	0.49	0.6879	1.00	0.8413
-0.52	0.3015	-0.01	0.4960	0.50	0.6915	1.01	0.8438

Table C.1 Standard normal cumulative distribution function, M (continued).

z	M(*z*)	*z*	M(*z*)	*z*	M(*z*)	*z*	M(*z*)
1.02	0.8461	1.52	0.9357	2.02	0.9783	2.52	0.9941
1.03	0.8485	1.53	0.9370	2.03	0.9788	2.53	0.9943
1.04	0.8508	1.54	0.9382	2.04	0.9793	2.54	0.9945
1.05	0.8531	1.55	0.9394	2.05	0.9798	2.55	0.9946
1.06	0.8554	1.56	0.9406	2.06	0.9803	2.56	0.9948
1.07	0.8577	1.57	0.9418	2.07	0.9808	2.57	0.9949
1.08	0.8599	1.58	0.9429	2.08	0.9812	2.58	0.9951
1.09	0.8621	1.59	0.9441	2.09	0.9817	2.59	0.9952
1.10	0.8643	1.60	0.9452	2.10	0.9821	2.60	0.9953
1.11	0.8665	1.61	0.9463	2.11	0.9826	2.61	0.9955
1.12	0.8686	1.62	0.9474	2.12	0.9830	2.62	0.9956
1.13	0.8708	1.63	0.9484	2.13	0.9834	2.63	0.9957
1.14	0.8729	1.64	0.9495	2.14	0.9838	2.64	0.9959
1.15	0.8749	1.65	0.9505	2.15	0.9842	2.65	0.9960
1.16	0.8770	1.66	0.9515	2.16	0.9846	2.66	0.9961
1.17	0.8790	1.67	0.9525	2.17	0.9850	2.67	0.9962
1.18	0.8810	1.68	0.9535	2.18	0.9854	2.68	0.9963
1.19	0.8830	1.69	0.9545	2.19	0.9857	2.69	0.9964
1.20	0.8849	1.70	0.9554	2.20	0.9861	2.70	0.9965
1.21	0.8869	1.71	0.9564	2.21	0.9864	2.71	0.9966
1.22	0.8888	1.72	0.9573	2.22	0.9868	2.72	0.9967
1.23	0.8907	1.73	0.9582	2.23	0.9871	2.73	0.9968
1.24	0.8925	1.74	0.9591	2.24	0.9875	2.74	0.9969
1.25	0.8944	1.75	0.9599	2.25	0.9878	2.75	0.9970
1.26	0.8962	1.76	0.9608	2.26	0.9881	2.76	0.9971
1.27	0.8980	1.77	0.9616	2.27	0.9884	2.77	0.9972
1.28	0.8997	1.78	0.9625	2.28	0.9887	2.78	0.9973
1.29	0.9015	1.79	0.9633	2.29	0.9890	2.79	0.9974
1.30	0.9032	1.80	0.9641	2.30	0.9893	2.80	0.9974
1.31	0.9049	1.81	0.9649	2.31	0.9896	2.81	0.9975
1.32	0.9066	1.82	0.9656	2.32	0.9898	2.82	0.9976
1.33	0.9082	1.83	0.9664	2.33	0.9901	2.83	0.9977
1.34	0.9099	1.84	0.9671	2.34	0.9904	2.84	0.9977
1.35	0.9115	1.85	0.9678	2.35	0.9906	2.85	0.9978
1.36	0.9131	1.86	0.9686	2.36	0.9909	2.86	0.9979
1.37	0.9147	1.87	0.9693	2.37	0.9911	2.87	0.9979
1.38	0.9162	1.88	0.9699	2.38	0.9913	2.88	0.9980
1.39	0.9177	1.89	0.9706	2.39	0.9916	2.89	0.9981
1.40	0.9192	1.90	0.9713	2.40	0.9918	2.90	0.9981
1.41	0.9207	1.91	0.9719	2.41	0.9920	2.91	0.9982
1.42	0.9222	1.92	0.9726	2.42	0.9922	2.92	0.9982
1.43	0.9236	1.93	0.9732	2.43	0.9925	2.93	0.9983
1.44	0.9251	1.94	0.9738	2.44	0.9927	2.94	0.9984
1.45	0.9265	1.95	0.9744	2.45	0.9929	2.95	0.9984
1.46	0.9279	1.96	0.9750	2.46	0.9931	2.96	0.9985
1.47	0.9292	1.97	0.9756	2.47	0.9932	2.97	0.9985
1.48	0.9306	1.98	0.9761	2.48	0.9934	2.98	0.9986
1.49	0.9319	1.99	0.9767	2.49	0.9936	2.99	0.9986
1.50	0.9332	2.00	0.9772	2.50	0.9938	3.00	0.9987
1.51	0.9345	2.01	0.9778	2.51	0.9940	3.50	0.99977
						4.00	0.999968

Table C.2. Percentiles of the chi-squared distribution.



Deg. of freedom	$P^2_{0.005}$	$P^2_{0.01}$	$P^2_{0.025}$	$P^2_{0.05}$	$P^2_{0.10}$	$P^2_{0.20}$	$P^2_{0.30}$	$P^2_{0.40}$
0.25	4.8E-19	1.2E-16	1.9E-13	4.8E-11	1.24E-8	3.17E-6	8.12E-5	8.11E-4
0.5	8.4E-10	1.35E-8	5.27E-7	8.44E-6	1.35E-4	2.16E-3	0.0110	0.0350
1	3.93E-5	1.57E-4	9.82E-4	3.93E-3	0.0158	0.0642	0.148	0.275
2	0.0100	0.0201	0.0506	0.103	0.211	0.446	0.713	1.022
3	0.0717	0.115	0.216	0.352	0.584	1.005	1.424	1.869
4	0.207	0.297	0.484	0.711	1.064	1.649	2.195	2.753
5	0.412	0.554	0.831	1.145	1.610	2.343	3.000	3.655
6	0.676	0.872	1.237	1.635	2.204	3.070	3.828	4.570
7	0.989	1.239	1.690	2.167	2.833	3.822	4.671	5.493
8	1.344	1.646	2.180	2.733	3.490	4.594	5.527	6.423
9	1.735	2.088	2.700	3.325	4.168	5.380	6.393	7.357
10	2.156	2.558	3.247	3.940	4.865	6.179	7.267	8.295
11	2.603	3.053	3.816	4.575	5.578	6.989	8.148	9.237
12	3.074	3.571	4.404	5.226	6.304	7.807	9.034	10.18
13	3.565	4.107	5.009	5.892	7.042	8.634	9.926	11.13
14	4.075	4.660	5.629	6.571	7.790	9.467	10.82	12.08
15	4.601	5.229	6.262	7.261	8.547	10.31	11.72	13.03
16	5.142	5.812	6.908	7.962	9.312	11.15	12.62	13.98
17	5.697	6.408	7.564	8.672	10.09	12.00	13.53	14.94
18	6.265	7.015	8.231	9.390	10.86	12.86	14.44	15.89
19	6.844	7.633	8.907	10.12	11.65	13.72	15.35	16.85
20	7.434	8.260	9.591	10.85	12.44	14.58	16.27	17.81
21	8.034	8.897	10.28	11.59	13.24	15.44	17.18	18.77
22	8.643	9.542	10.98	12.34	14.04	16.31	18.10	19.73
23	9.260	10.20	11.69	13.09	14.85	17.19	19.02	20.69
24	9.886	10.86	12.40	13.85	15.66	18.06	19.94	21.65
25	10.52	11.52	13.12	14.61	16.47	18.94	20.87	22.62
26	11.16	12.20	13.84	15.38	17.29	19.82	21.79	23.58
27	11.81	12.88	14.57	16.15	18.11	20.70	22.72	24.54
28	12.46	13.56	15.31	16.93	18.94	21.59	23.65	25.51
29	13.12	14.26	16.05	17.71	19.77	22.48	24.58	26.48
30	13.79	14.95	16.79	18.49	20.60	23.36	25.51	27.44
35	17.19	18.51	20.57	22.47	24.80	27.84	30.18	32.28
40	20.71	22.16	24.43	26.51	29.05	32.34	34.87	37.13
45	24.31	25.90	28.37	30.61	33.35	36.88	39.58	42.00
50	27.99	29.71	32.36	34.76	37.69	41.45	44.31	46.86
55	31.73	33.57	36.40	38.96	42.06	46.04	49.06	51.74
60	35.53	37.48	40.48	43.19	46.46	50.64	53.81	56.62
70	43.25	45.42	48.75	51.74	55.33	59.90	63.35	66.40
80	51.14	53.52	57.15	60.39	64.28	69.21	72.92	76.19
90	59.17	61.74	65.64	69.13	73.29	78.56	82.52	85.60
100	67.30	70.05	74.22	77.93	82.36	87.95	92.13	95.81
125	88.01	91.17	95.94	100.2	105.2	111.5	116.3	120.4
150	109.1	112.7	118.0	122.7	128.3	135.3	140.5	145.0

For large degrees of freedom *d*, use $\chi^2_p \approx \left(z_p + \sqrt{2d-1}\right)^2 / 2$, where z_p is the corresponding percentile of a standard normal distribution.

Table C.2 Percentiles of the chi-squared distribution (continued).

$P^2_{0.50}$	$P^2_{0.60}$	$P^2_{0.70}$	$P^2_{0.80}$	$P^2_{0.90}$	$P^2_{0.95}$	$P^2_{0.975}$	$P^2_{0.99}$	$P^2_{0.995}$	Deg. of freedom
4.84E-3	0.0210	0.0737	0.229	0.716	1.419	2.269	3.543	4.585	0.25
0.0873	0.188	0.375	0.726	1.501	2.420	3.433	4.868	6.004	0.5
0.455	0.708	1.074	1.642	2.706	3.841	5.024	6.635	7.879	1
1.386	1.833	2.408	3.219	4.605	5.991	7.378	9.210	10.56	2
2.366	2.946	3.665	4.642	6.251	7.815	9.348	11.34	12.84	3
3.357	4.045	4.878	5.989	7.779	9.488	11.14	13.28	14.86	4
4.351	5.132	6.064	7.289	9.236	11.07	12.83	15.09	16.75	5
5.348	6.211	7.231	8.558	10.64	12.59	14.45	16.81	18.55	6
6.346	7.283	8.383	9.803	12.02	14.07	16.01	18.48	20.28	7
7.344	8.351	9.524	11.03	13.36	15.51	17.53	20.09	21.95	8
8.343	9.414	10.66	12.24	14.68	16.92	19.02	21.67	23.59	9
9.342	10.47	11.78	13.44	15.99	18.31	20.48	23.21	25.19	10
10.34	11.53	12.90	14.63	17.28	19.68	21.92	24.72	26.76	11
11.34	12.58	14.01	15.81	18.55	21.03	23.34	26.22	28.30	12
12.34	13.64	15.12	16.98	19.81	22.36	24.74	27.69	29.82	13
13.34	14.69	16.22	18.15	21.06	23.68	26.12	29.14	31.32	14
14.34	15.73	17.32	19.31	22.31	25.00	27.49	30.58	32.80	15
15.34	16.78	18.42	20.47	23.54	26.30	28.85	32.00	34.27	16
16.34	17.82	19.51	21.61	24.77	27.59	30.19	33.41	35.72	17
17.34	18.87	20.60	22.76	25.99	28.87	31.53	34.81	37.16	18
18.34	19.91	21.69	23.90	27.20	30.14	32.85	36.19	38.58	19
19.34	20.95	22.77	25.04	28.41	31.41	34.17	37.57	40.00	20
20.34	21.99	23.86	26.17	29.62	32.67	35.48	38.93	41.40	21
21.34	23.03	24.94	27.30	30.81	33.92	36.78	40.29	42.80	22
22.34	24.07	26.02	28.43	32.01	35.17	38.08	41.64	44.18	23
23.34	25.11	27.10	29.55	33.20	36.42	39.36	42.98	45.56	24
24.34	26.14	28.17	30.68	34.38	37.65	40.65	44.31	46.93	25
25.34	27.18	29.25	31.79	35.56	38.89	41.92	45.64	48.29	26
26.34	28.21	30.32	32.91	36.74	40.11	43.19	46.96	49.64	27
27.34	29.25	31.39	34.03	37.92	41.34	44.46	48.28	50.99	28
28.34	30.28	32.46	35.14	39.09	42.56	45.72	49.59	52.34	29
29.34	31.32	33.53	36.25	40.26	43.77	46.98	50.89	53.67	30
34.34	36.47	38.86	41.78	46.06	49.80	53.20	57.34	60.27	35
39.34	41.62	44.16	47.27	51.81	55.76	59.34	63.69	66.77	40
44.34	46.76	49.45	52.73	57.51	61.66	65.41	69.96	73.17	45
49.33	51.89	54.72	58.16	63.17	67.50	71.42	76.15	79.49	50
54.33	57.02	59.98	63.58	68.80	73.31	77.38	82.29	85.75	55
59.33	62.13	65.23	68.97	74.40	79.08	83.30	88.38	91.95	60
69.34	72.36	75.69	79.71	85.52	90.53	95.03	100.4	104.2	70
79.34	82.56	86.12	90.40	96.57	101.9	106.6	112.3	116.3	80
89.33	92.76	96.52	101.0	107.6	113.1	118.1	124.1	128.3	90
99.33	102.9	106.9	111.7	118.5	124.3	129.6	135.8	140.2	100
124.3	128.4	132.8	138.1	145.6	152.1	157.8	164.7	169.5	125
149.3	153.8	158.6	164.3	172.6	179.6	185.8	193.2	198.4	150

Table C.3. Percentiles of Student's t distribution.

Deg. of freedom	$t_{0.6}$	$t_{0.7}$	$t_{0.8}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	0.325	0.727	1.376	3.078	6.314	12.71	31.82	63.66
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
50	0.255	0.528	0.849	1.299	1.676	2.009	2.403	2.678
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
70	0.254	0.527	0.847	1.294	1.667	1.994	2.381	2.648
80	0.254	0.526	0.846	1.292	1.664	1.990	2.374	2.639
90	0.254	0.526	0.846	1.291	1.662	1.987	2.368	2.632
100	0.254	0.526	0.845	1.290	1.660	1.984	2.364	2.626
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
150	0.254	0.526	0.844	1.287	1.655	1.976	2.351	2.609
4	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576

For percentiles below the 50th, use the fact that the *t* distribution is symmetrical about zero, so $t_{1!q} = !t_q$.

Tables of the beta(α, β) distribution are given on the next pages. Because of the limited size of the paper, each table has been split into two pieces and printed on two facing pages. Each table contains a diagonal line, shown by blackened cells. The table entries below this line and to the left are the lower percentiles, such as the 10th or the 5th. The table entries above the diagonal and to the right are the upper percentiles, such as the 90th or the 95th. In this way, both sets of percentiles appear in a single table.

Only distributions with $\alpha < \beta$ are tabulated. These distributions have probability concentrated near zero, and are the distributions usually encountered in PRA work. For distributions with $\alpha > \beta$, use the fact that if X has a beta(α, β) distribution then $1 - X$ has a beta(β, α) distribution. An example is given as a footnote to each table.

The size of the page limits the number of parameter pairs (α, β) that can be tabulated. Therefore, interpolation is often necessary, which may give only rough accuracy. If greater accuracy is required, the user can find the beta distribution calculated by many commonly used computer packages and spreadsheets. Similarly, extrapolation beyond the table may sometimes be necessary. A footnote to each table gives an approximate extrapolation formula when $\beta >> \alpha$, and an example of its application. If greater accuracy is needed, use a commercially available computer program.

Table C.4. 90th and 10th percentiles of beta (″, $) distribution.

		$ for 90th percentiles												
		0.1	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0		
		1.00E+0	7.33E-1	3.49E-1	2.16E-1	1.55E-1	1.20E-1	9.84E-2	8.31E-2	7.20E-2	6.34E-2	5.67E-2	0.1	
		[redacted]	9.76E-1	8.10E-1	6.49E-1	5.32E-1	4.48E-1	3.86E-1	3.39E-1	3.02E-1	2.72E-1	2.47E-1	0.5	
	0.1	8.87E-8	[redacted]	9.00E-1	7.85E-1	6.84E-1	6.02E-1	5.36E-1	4.82E-1	4.38E-1	4.01E-1	3.69E-1	1.0	
	0.5	3.46E-10	2.45E-2	[redacted]	8.44E-1	7.59E-1	6.85E-1	6.22E-1	5.68E-1	5.23E-1	4.84E-1	4.50E-1	1.5	
	1.0	1.00E-10	1.00E-2	1.00E-1	[redacted]	8.04E-1	7.38E-1	6.80E-1	6.28E-1	5.84E-1	5.45E-1	5.10E-1	2.0	
	1.5	5.60E-11	6.18E-3	6.78E-2	1.56E-1	[redacted]	7.75E-1	7.21E-1	6.73E-1	6.30E-1	5.92E-1	5.58E-1	2.5	″
	2.0	3.86E-11	4.46E-3	5.13E-2	1.23E-1	1.96E-1	[redacted]	7.53E-1	7.08E-1	6.67E-1	6.30E-1	5.96E-1	3.0	for
	2.5	2.93E-11	3.48E-3	4.13E-2	1.02E-1	1.65E-1	2.25E-1	[redacted]	7.36E-1	6.97E-1	6.61E-1	6.28E-1	3.5	90th
	3.0	2.37E-11	2.86E-3	3.45E-2	8.64E-2	1.43E-1	1.97E-1	2.47E-1	[redacted]	7.21E-1	6.87E-1	6.55E-1	4.0	%iles
	3.5	1.98E-11	2.42E-3	2.97E-2	7.53E-2	1.26E-1	1.75E-1	2.21E-1	2.64E-1	[redacted]	7.09E-1	6.79E-1	4.5	
	4.0	1.71E-11	2.10E-3	2.60E-2	6.66E-2	1.12E-1	1.58E-1	2.01E-1	2.41E-1	2.79E-1	[redacted]	6.99E-1	5.0	
	4.5	1.50E-11	1.85E-3	2.31E-2	5.98E-2	1.01E-1	1.43E-1	1.84E-1	2.22E-1	2.58E-1	2.91E-1	[redacted]		
$ for 10th %iles	5.	1.33E-11	1.66E-3	2.09E-2	5.42E-2	9.26E-2	1.32E-1	1.70E-1	2.06E-1	2.40E-1	2.71E-1	3.01E-1		
	6.	1.09E-11	1.37E-3	1.74E-2	4.57E-2	7.88E-2	1.13E-1	1.47E-1	1.79E-1	2.10E-1	2.40E-1	2.67E-1		
	7.	9.26E-12	1.17E-3	1.49E-2	3.95E-2	6.86E-2	9.91E-2	1.29E-1	1.59E-1	1.88E-1	2.15E-1	2.41E-1		
	8.	8.04E-12	1.02E-3	1.31E-2	3.48E-2	6.08E-2	8.82E-2	1.16E-1	1.43E-1	1.69E-1	1.95E-1	2.19E-1		
	9.	7.10E-12	9.02E-4	1.16E-2	3.11E-2	5.45E-2	7.95E-2	1.05E-1	1.30E-1	1.54E-1	1.78E-1	2.01E-1		
	10.	6.36E-12	8.09E-4	1.05E-2	2.81E-2	4.95E-2	7.23E-2	9.57E-2	1.19E-1	1.42E-1	1.64E-1	1.85E-1		
	12.5	5.04E-12	6.44E-4	8.39E-3	2.27E-2	4.01E-2	5.90E-2	7.86E-2	9.82E-2	1.18E-1	1.37E-1	1.55E-1		
	15.	4.17E-12	5.35E-4	7.00E-3	1.90E-2	3.37E-2	4.99E-2	6.67E-2	8.37E-2	1.01E-1	1.17E-1	1.34E-1		
	20.	3.11E-12	4.00E-4	5.25E-3	1.43E-2	2.56E-2	3.81E-2	5.12E-2	6.46E-2	7.81E-2	9.16E-2	1.05E-1		
	30.	2.06E-12	2.65E-4	3.51E-3	9.61E-3	1.73E-2	2.59E-2	3.49E-2	4.43E-2	5.39E-2	6.36E-2	7.33E-2		
	50.	1.23E-12	1.59E-4	2.10E-3	5.80E-3	1.05E-2	1.57E-2	2.14E-2	2.73E-2	3.33E-2	3.95E-2	4.57E-2		
	100.	6.10E-13	7.91E-5	1.05E-3	2.91E-3	5.28E-3	7.96E-3	1.09E-2	1.39E-2	1.70E-2	2.03E-2	2.36E-2		
		0.1	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0		
		″ for 10th percentiles												

For example, the 10th percentile of a beta(1.5, 10) distribution is 2.81E!2. The 90th percentile of the same distribution (from the table continuation on the next page) is 2.63E!1. If X has a beta(″, $) distributions with " > $, use the relation $X = 1 ! Y$, where Y has a beta($, ″) distribution. Thus, for example, the 10th and 90th percentiles of a beta(10, 1.5) distribution are 1 ! 2.63E!1 = 0.737 and 1 ! 2.81E!2 = 0.9719, respectively.

For a beta(″, $) distribution with $ >> ", the p quantile is approximated by $\chi^2_p(2\alpha) / [2\beta + \chi^2_p(2\alpha)]$. For example, the 10th percentile of a beta(2,100) distribution, shown above as 5.28E!3, is approximated by $\chi^2_{0.10}(4) / [200 + \chi^2_{0.10}(4)] = 1.064/201.064 = 5.29E!3$. The 90th percentile, 3.80E!2, is approximated by 7.770/207.779 = 3.74E!2.

Table C.4 90th and 10th percentiles of beta (″, $) distribution (continued).

		$ for 90th percentiles												
		6.	7.	8.	9.	10.	12.5	15.	20.	30.	50.	100.		
		4.67E-2	3.98E-2	3.46E-2	3.06E-2	2.75E-2	2.18E-2	1.81E-2	1.35E-2	8.97E-3	5.36E-3	2.67E-3	0.1	″ for 90th %iles
		2.09E-1	1.81E-1	1.60E-1	1.43E-1	1.29E-1	1.05E-1	8.76E-2	6.62E-2	4.45E-2	2.68E-2	1.35E-2	0.5	
		3.19E-1	2.80E-1	2.50E-1	2.26E-1	2.06E-1	1.68E-1	1.42E-1	1.09E-1	7.39E-2	4.50E-2	2.28E-2	1.0	
		3.94E-1	3.51E-1	3.16E-1	2.87E-1	2.63E-1	2.18E-1	1.85E-1	1.43E-1	9.82E-2	6.03E-2	3.07E-2	1.5	
		4.53E-1	4.06E-1	3.68E-1	3.37E-1	3.10E-1	2.59E-1	2.22E-1	1.73E-1	1.20E-1	7.41E-2	3.80E-2	2.0	
		4.99E-1	4.52E-1	4.12E-1	3.79E-1	3.50E-1	2.95E-1	2.55E-1	2.00E-1	1.40E-1	8.70E-2	4.48E-2	2.5	
		5.38E-1	4.90E-1	4.50E-1	4.15E-1	3.86E-1	3.27E-1	2.84E-1	2.24E-1	1.58E-1	9.91E-2	5.13E-2	3.0	
		5.71E-1	5.23E-1	4.82E-1	4.47E-1	4.17E-1	3.56E-1	3.10E-1	2.47E-1	1.75E-1	1.11E-1	5.76E-2	3.5	
		5.99E-1	5.52E-1	5.11E-1	4.75E-1	4.44E-1	3.82E-1	3.34E-1	2.68E-1	1.91E-1	1.22E-1	6.37E-2	4.0	
		6.24E-1	5.77E-1	5.36E-1	5.01E-1	4.69E-1	4.06E-1	3.57E-1	2.87E-1	2.07E-1	1.32E-1	6.96E-2	4.5	
		6.46E-1	5.99E-1	5.59E-1	5.23E-1	4.92E-1	4.27E-1	3.78E-1	3.06E-1	2.21E-1	1.43E-1	7.54E-2	5.	
		6.82E-1	6.38E-1	5.98E-1	5.63E-1	5.32E-1	4.66E-1	4.15E-1	3.40E-1	2.49E-1	1.62E-1	8.65E-2	6.	
		[redacted]	6.69E-1	6.31E-1	5.96E-1	5.65E-1	5.00E-1	4.48E-1	3.70E-1	2.74E-1	1.80E-1	9.72E-2	7.	
$ for 10th %iles	6.	3.18E-1	[redacted]	6.58E-1	6.25E-1	5.94E-1	5.29E-1	4.77E-1	3.97E-1	2.98E-1	1.98E-1	1.08E-1	8.	
	7.	2.88E-1	3.31E-1	[redacted]	6.50E-1	6.20E-1	5.56E-1	5.03E-1	4.22E-1	3.19E-1	2.14E-1	1.18E-1	9.	
	8.	2.64E-1	3.05E-1	3.42E-1	[redacted]	6.42E-1	5.79E-1	5.26E-1	4.45E-1	3.40E-1	2.30E-1	1.27E-1	10.	
	9.	2.43E-1	2.82E-1	3.18E-1	3.50E-1	[redacted]	6.27E-1	5.76E-1	4.95E-1	3.85E-1	2.66E-1	1.50E-1	12.5	
	10.	2.26E-1	2.63E-1	2.97E-1	3.29E-1	3.58E-1	[redacted]	6.16E-1	5.36E-1	4.25E-1	2.99E-1	1.72E-1	15.	
	12.5	1.91E-1	2.25E-1	2.56E-1	2.85E-1	3.12E-1	3.73E-1	[redacted]	6.01E-1	4.89E-1	3.56E-1	2.11E-1	20.	
	15.	1.66E-1	1.96E-1	2.25E-1	2.52E-1	2.77E-1	3.34E-1	3.84E-1	[redacted]	5.82E-1	4.45E-1	2.79E-1	30.	
	20.	1.31E-1	1.57E-1	1.81E-1	2.04E-1	2.26E-1	2.78E-1	3.23E-1	3.99E-1	[redacted]	5.64E-1	3.83E-1	50.	
	30.	9.26E-2	1.12E-1	1.30E-1	1.48E-1	1.66E-1	2.07E-1	2.45E-1	3.12E-1	4.18E-1	[redacted]	5.45E-1	100.	
	50.	5.83E-2	7.09E-2	8.35E-2	9.59E-2	1.08E-1	1.38E-1	1.66E-1	2.18E-1	3.06E-1	4.36E-1	[redacted]		
	100.	3.03E-2	3.71E-2	4.40E-2	5.09E-2	5.78E-2	7.50E-2	9.19E-2	1.25E-1	1.84E-1	2.85E-1	4.55E-1		
		6.	7.	8.	9.	10.	12.5	15.	20.	30.	50.	100.		
		″ for 10th percentiles												

Table C.5. 95th and 5th percentiles of beta (″, $) distribution.

$ for 95th percentiles												
0.1	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0		
1.00E+0	9.24E-1	5.99E-1	4.09E-1	3.06E-1	2.44E-1	2.02E-1	1.72E-1	1.50E-1	1.33E-1	1.19E-1	0.1	
[redacted]	9.94E-1	9.02E-1	7.71E-1	6.58E-1	5.69E-1	4.99E-1	4.44E-1	3.99E-1	3.62E-1	3.32E-1	0.5	
0.1	8.66E-11	[redacted]	9.50E-1	8.64E-1	7.76E-1	6.98E-1	6.32E-1	5.75E-1	5.27E-1	4.86E-1	4.51E-1	1.0
0.5	3.38E-13	6.16E-3	[redacted]	9.03E-1	8.32E-1	7.64E-1	7.04E-1	6.51E-1	6.04E-1	5.63E-1	5.27E-1	1.5
1.0	9.77E-14	2.50E-3	5.00E-2	[redacted]	8.65E-1	8.06E-1	7.51E-1	7.02E-1	6.57E-1	6.18E-1	5.82E-1	2.0
1.5	5.46E-14	1.54E-3	3.36E-2	9.73E-2	[redacted]	8.35E-1	7.85E-1	7.39E-1	6.97E-1	6.59E-1	6.24E-1	2.5
2.0	3.77E-14	1.11E-3	2.53E-2	7.60E-2	1.35E-1	[redacted]	8.11E-1	7.68E-1	7.29E-1	6.92E-1	6.59E-1	3.0
2.5	2.87E-14	8.68E-4	2.03E-2	6.24E-2	1.13E-1	1.65E-1	[redacted]	7.91E-1	7.54E-1	7.19E-1	6.87E-1	3.5
3.0	2.31E-14	7.12E-4	1.70E-2	5.30E-2	9.76E-2	1.44E-1	1.89E-1	[redacted]	7.75E-1	7.42E-1	7.11E-1	4.0
3.5	1.94E-14	6.03E-4	1.45E-2	4.60E-2	8.57E-2	1.28E-1	1.69E-1	2.09E-1	[redacted]	7.61E-1	7.31E-1	4.5
4.0	1.67E-14	5.23E-4	1.27E-2	4.07E-2	7.64E-2	1.15E-1	1.53E-1	1.90E-1	2.25E-1	[redacted]	7.49E-1	5.0
4.5	1.46E-14	4.62E-4	1.13E-2	3.64E-2	6.90E-2	1.04E-1	1.40E-1	1.75E-1	2.08E-1	2.39E-1	[redacted]	
5.	1.30E-14	4.13E-4	1.02E-2	3.30E-2	6.28E-2	9.55E-2	1.29E-1	1.61E-1	1.93E-1	2.23E-1	2.51E-1	
6.	1.07E-14	3.42E-4	8.51E-3	2.78E-2	5.34E-2	8.18E-2	1.11E-1	1.40E-1	1.69E-1	1.96E-1	2.22E-1	
7.	9.05E-15	2.91E-4	7.30E-3	2.40E-2	4.64E-2	7.15E-2	9.77E-2	1.24E-1	1.50E-1	1.75E-1	2.00E-1	
8.	7.85E-15	2.54E-4	6.39E-3	2.11E-2	4.10E-2	6.36E-2	8.73E-2	1.11E-1	1.35E-1	1.58E-1	1.81E-1	
9.	6.93E-15	2.25E-4	5.68E-3	1.89E-2	3.68E-2	5.72E-2	7.88E-2	1.01E-1	1.23E-1	1.45E-1	1.66E-1	
10.	6.21E-15	2.02E-4	5.12E-3	1.70E-2	3.33E-2	5.20E-2	7.19E-2	9.22E-2	1.13E-1	1.33E-1	1.53E-1	
12.5	4.92E-15	1.60E-4	4.10E-3	1.37E-2	2.70E-2	4.24E-2	5.89E-2	7.60E-2	9.33E-2	1.11E-1	1.28E-1	
15.	4.08E-15	1.33E-4	3.41E-3	1.15E-2	2.27E-2	3.57E-2	4.99E-2	6.47E-2	7.97E-2	9.48E-2	1.10E-1	
20.	3.03E-15	9.95E-5	2.56E-3	8.65E-3	1.72E-2	2.72E-2	3.82E-2	4.98E-2	6.17E-2	7.37E-2	8.59E-2	
30.	2.01E-15	6.61E-5	1.71E-3	5.80E-3	1.16E-2	1.85E-2	2.60E-2	3.41E-2	4.25E-2	5.11E-2	5.98E-2	
50.	1.20E-15	3.95E-5	1.03E-3	3.49E-3	7.01E-3	1.12E-2	1.59E-2	2.09E-2	2.62E-2	3.16E-2	3.72E-2	
100.	5.96E-16	1.97E-5	5.13E-4	1.75E-3	3.53E-3	5.67E-3	8.06E-3	1.06E-2	1.34E-2	1.62E-2	1.91E-2	
0.1	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0		
″ for 5th percentiles												

$ for 5th %iles

″ for 95th %iles

For example, the 5th percentile of a beta(1.5, 10) distribution is 1.70E!2. The 95th percentile of the same distribution (from the table continuation on the next page) is 3.17E!1. If X has a beta(″, $) distributions with " > $, use the relation $X = 1 ! Y$, where Y has a beta($, ″) distribution. Thus, for example, the 5th and 95th percentiles of a beta(10, 1.5) distribution are 1 ! 3.17E!1 = 0.683 and 1 ! 1.70E!2 = 0.9830, respectively.

For a beta(″, $) distribution with $>> ″, the p quantile is approximated by $\chi^2_p(2\alpha) / [2\beta + \chi^2_p(2\alpha)]$. For example, the 5th percentile of a beta(2,100) distribution, shown here as 3.53E!3, is approximated by $\chi^2_{0.05}(4) / [200 + \chi^2_{0.05}(4)] = 0.711/200.711 = 3.54E!3$. The 95th percentile, 4.61E!2, is approximated by 9.488/209.488 = 4.53E!2.

Table C.5 95th and 5th percentiles of beta (″, $) distribution (continued).

		$ for 95th percentiles											
		6.	7.	8.	9.	10.	12.5	15.	20.	30.	50.	100.	
		9.91E-2	8.47E-2	7.39E-2	6.56E-2	5.89E-2	4.70E-2	3.91E-2	2.92E-2	1.94E-2	1.16E-2	5.81E-3	0.1
		2.83E-1	2.47E-1	2.19E-1	1.97E-1	1.79E-1	1.45E-1	1.22E-1	9.27E-2	6.25E-2	3.79E-2	1.91E-2	0.5
		3.93E-1	3.48E-1	3.12E-1	2.83E-1	2.59E-1	2.13E-1	1.81E-1	1.39E-1	9.50E-2	5.82E-2	2.95E-2	1.0
		4.66E-1	4.17E-1	3.78E-1	3.45E-1	3.17E-1	2.64E-1	2.26E-1	1.76E-1	1.21E-1	7.48E-2	3.82E-2	1.5
		5.21E-1	4.71E-1	4.29E-1	3.94E-1	3.64E-1	3.06E-1	2.64E-1	2.07E-1	1.44E-1	8.97E-2	4.61E-2	2.0
		5.64E-1	5.14E-1	4.71E-1	4.35E-1	4.04E-1	3.42E-1	2.97E-1	2.34E-1	1.65E-1	1.03E-1	5.35E-2	2.5
		6.00E-1	5.50E-1	5.07E-1	4.70E-1	4.38E-1	3.74E-1	3.26E-1	2.59E-1	1.84E-1	1.16E-1	6.04E-2	3.0
		6.30E-1	5.80E-1	5.38E-1	5.01E-1	4.68E-1	4.02E-1	3.53E-1	2.82E-1	2.02E-1	1.28E-1	6.71E-2	3.5
		6.55E-1	6.07E-1	5.64E-1	5.27E-1	4.95E-1	4.28E-1	3.77E-1	3.04E-1	2.19E-1	1.40E-1	7.36E-2	4.0
		6.77E-1	6.30E-1	5.88E-1	5.51E-1	5.18E-1	4.51E-1	3.99E-1	3.23E-1	2.34E-1	1.51E-1	7.98E-2	4.5
		6.96E-1	6.50E-1	6.09E-1	5.73E-1	5.40E-1	4.72E-1	4.19E-1	3.42E-1	2.49E-1	1.62E-1	8.59E-2	5.
		7.29E-1	6.85E-1	6.45E-1	6.10E-1	5.77E-1	5.10E-1	4.56E-1	3.75E-1	2.77E-1	1.82E-1	9.75E-2	6.
		[redacted]	7.13E-1	6.75E-1	6.40E-1	6.09E-1	5.42E-1	4.87E-1	4.05E-1	3.03E-1	2.01E-1	1.09E-1	7.
$ for 5th %iles	6.	2.71E-1	[redacted]	7.00E-1	6.67E-1	6.36E-1	5.70E-1	5.15E-1	4.32E-1	3.26E-1	2.18E-1	1.19E-1	8.
	7.	2.45E-1	2.87E-1	[redacted]	6.89E-1	6.59E-1	5.94E-1	5.40E-1	4.57E-1	3.48E-1	2.35E-1	1.30E-1	9.
	8.	2.24E-1	2.64E-1	3.00E-1	[redacted]	6.80E-1	6.16E-1	5.63E-1	4.79E-1	3.68E-1	2.51E-1	1.40E-1	10.
	9.	2.06E-1	2.44E-1	2.79E-1	3.11E-1	[redacted]	6.62E-1	6.10E-1	5.27E-1	4.13E-1	2.88E-1	1.63E-1	12.5
	10.	1.91E-1	2.27E-1	2.60E-1	2.91E-1	3.20E-1	[redacted]	6.48E-1	5.67E-1	4.52E-1	3.21E-1	1.85E-1	15.
	12.5	1.61E-1	1.93E-1	2.23E-1	2.52E-1	2.78E-1	3.38E-1	[redacted]	6.29E-1	5.15E-1	3.77E-1	2.25E-1	20.
	15.	1.40E-1	1.68E-1	1.96E-1	2.22E-1	2.46E-1	3.03E-1	3.52E-1	[redacted]	6.05E-1	4.65E-1	2.94E-1	30.
	20.	1.10E-1	1.34E-1	1.57E-1	1.79E-1	2.00E-1	2.50E-1	2.95E-1	3.71E-1	[redacted]	5.82E-1	3.98E-1	50.
	30.	7.74E-2	9.50E-2	1.12E-1	1.29E-1	1.46E-1	1.86E-1	2.23E-1	2.89E-1	3.95E-1	[redacted]	5.58E-1	100.
	50.	4.86E-2	6.02E-2	7.18E-2	8.34E-2	9.49E-2	1.23E-1	1.50E-1	2.01E-1	2.88E-1	4.18E-1	[redacted]	
	100.	2.52E-2	3.14E-2	3.77E-2	4.42E-2	5.06E-2	6.68E-2	8.29E-2	1.14E-1	1.73E-1	2.71E-1	4.42E-1	
		6.	7.	8.	9.	10.	12.5	15.	20.	30.	50.	100.	
		″ for 5th percentiles											

″ for 95th %iles

Table C.6. 97.5th and 2.5th percentiles of beta (″, $) distribution.

		$ for 97.5th percentiles											
		0.1	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	
		1.00E+0	9.80E-1	7.76E-1	5.84E-1	4.58E-1	3.74E-1	3.16E-1	2.72E-1	2.40E-1	2.14E-1	1.93E-1	0.1
		[redacted]	9.98E-1	9.51E-1	8.53E-1	7.53E-1	6.67E-1	5.95E-1	5.36E-1	4.86E-1	4.45E-1	4.10E-1	0.5
	0.1	8.46E-14	[redacted]	9.75E-1	9.15E-1	8.42E-1	7.71E-1	7.08E-1	6.51E-1	6.02E-1	5.59E-1	5.22E-1	1.0
	0.5	3.30E-16	1.54E-3	[redacted]	9.39E-1	8.82E-1	8.23E-1	7.67E-1	7.16E-1	6.70E-1	6.29E-1	5.91E-1	1.5
	1.0	9.54E-17	6.25E-4	2.50E-2	[redacted]	9.06E-1	8.55E-1	8.06E-1	7.59E-1	7.16E-1	6.77E-1	6.41E-1	2.0
	1.5	5.34E-17	3.86E-4	1.67E-2	6.08E-2	[redacted]	8.77E-1	8.33E-1	7.91E-1	7.51E-1	7.14E-1	6.79E-1	2.5
	2.0	3.68E-17	2.78E-4	1.26E-2	4.73E-2	9.43E-2	[redacted]	8.53E-1	8.14E-1	7.77E-1	7.42E-1	7.10E-1	3.0
	2.5	2.80E-17	2.17E-4	1.01E-2	3.87E-2	7.87E-2	1.23E-1	[redacted]	8.33E-1	7.98E-1	7.65E-1	7.34E-1	3.5
	3.0	2.26E-17	1.78E-4	8.40E-3	3.28E-2	6.76E-2	1.07E-1	1.47E-1	[redacted]	8.16E-1	7.85E-1	7.55E-1	4.0
	3.5	1.89E-17	1.51E-4	7.21E-3	2.85E-2	5.92E-2	9.44E-2	1.31E-1	1.67E-1	[redacted]	8.01E-1	7.73E-1	4.5
	4.0	1.63E-17	1.31E-4	6.31E-3	2.51E-2	5.27E-2	8.47E-2	1.18E-1	1.52E-1	1.84E-1	[redacted]	7.88E-1	5.0
$ for 2.5th %iles	4.5	1.43E-17	1.15E-4	5.61E-3	2.25E-2	4.75E-2	7.68E-2	1.08E-1	1.39E-1	1.69E-1	1.99E-1	[redacted]	
	5.	1.27E-17	1.03E-4	5.05E-3	2.04E-2	4.33E-2	7.02E-2	9.90E-2	1.28E-1	1.57E-1	1.85E-1	2.12E-1	
	6.	1.04E-17	8.53E-5	4.21E-3	1.71E-2	3.67E-2	6.00E-2	8.52E-2	1.11E-1	1.37E-1	1.62E-1	1.87E-1	
	7.	8.83E-18	7.27E-5	3.61E-3	1.48E-2	3.19E-2	5.24E-2	7.49E-2	9.81E-2	1.22E-1	1.45E-1	1.67E-1	
	8.	7.67E-18	6.33E-5	3.16E-3	1.30E-2	2.81E-2	4.65E-2	6.67E-2	8.78E-2	1.09E-1	1.31E-1	1.52E-1	
	9.	6.77E-18	5.61E-5	2.81E-3	1.16E-2	2.52E-2	4.18E-2	6.02E-2	7.95E-2	9.92E-2	1.19E-1	1.39E-1	
	10.	6.06E-18	5.03E-5	2.53E-3	1.05E-2	2.28E-2	3.80E-2	5.49E-2	7.27E-2	9.09E-2	1.09E-1	1.28E-1	
	12.5	4.81E-18	4.01E-5	2.02E-3	8.43E-3	1.85E-2	3.09E-2	4.49E-2	5.98E-2	7.52E-2	9.08E-2	1.07E-1	
	15.	3.98E-18	3.33E-5	1.69E-3	7.05E-3	1.55E-2	2.61E-2	3.80E-2	5.08E-2	6.41E-2	7.77E-2	9.15E-2	
	20.	2.96E-18[a]	2.49E-5	1.27E-3	5.31E-3	1.17E-2	1.98E-2	2.91E-2	3.90E-2	4.95E-2	6.03E-2	7.13E-2	
	30.	1.96E-18	1.65E-5	8.44E-4	3.56E-3	7.91E-3	1.34E-2	1.98E-2	2.67E-2	3.40E-2	4.17E-2	4.95E-2	
	50.	1.17E-18	9.87E-6	5.06E-4	2.14E-3	4.78E-3	8.16E-3	1.21E-2	1.64E-2	2.09E-2	2.58E-2	3.08E-2	
	100.	5.82E-19[a]	4.92E-6	2.53E-4	1.08E-3	2.41E-3	4.12E-3	6.11E-3	8.31E-3	1.07E-2	1.32E-2	1.58E-2	
		0.1	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	
		″ for 2.5th percentiles											

″ for 97.5th %iles

a. May be inaccurate. Calculation had not converged after 100 iterations.

For example, the 2.5th percentile of a beta(1.5, 10) distribution is 1.05E!2. The 97.5th percentile of the same distribution (from the table continuation on the next page) is 3.67E!1. If X has a beta(″, $) distributions with ″ > $, use the relation $X = 1 ! Y$, where Y has a beta($, ″) distribution. Thus, for example, the 2.5th and 97.5th percentiles of a beta(10, 1.5) distribution are 1 ! 3.67E!1 = 0.633 and 1 ! 1.05E!2 = 0.9895, respectively.

For a beta(″, $) distribution with $>> ", the p quantile is approximated by $\chi^2_p(2\alpha) / [2\beta + \chi^2_p(2\alpha)]$. For example, the 2.5th percentile of a beta(2,100) distribution, shown here as 2.41E!3, is approximated by $\chi^2_{0.025}(4) / [200 + \chi^2_{0.025}(4)]$ = 0.484/200.484 = 2.41E!3. The 97.5th percentile, 5.39E!2, is approximated by 11.14/211.14 = 5.28E!2.

Table C.6. 97.5th and 2.5th percentiles of beta (″, $) distribution (continued).

		$ for 97.5th percentiles											
		6.	7.	8.	9.	10.	12.5	15.	20.	30.	50.	100.	
		1.61E-1	1.38E-1	1.21E-1	1.08E-1	9.73E-2	7.79E-2	6.50E-2	4.88E-2	3.25E-2	1.95E-2	9.78E-3	0.1
		3.53E-1	3.10E-1	2.77E-1	2.49E-1	2.27E-1	1.85E-1	1.57E-1	1.19E-1	8.10E-2	4.92E-2	2.49E-2	0.5
		4.59E-1	4.10E-1	3.69E-1	3.36E-1	3.08E-1	2.56E-1	2.18E-1	1.68E-1	1.16E-1	7.11E-2	3.62E-2	1.0
		5.28E-1	4.76E-1	4.33E-1	3.97E-1	3.67E-1	3.07E-1	2.64E-1	2.06E-1	1.43E-1	8.88E-2	4.56E-2	1.5
		5.79E-1	5.27E-1	4.82E-1	4.45E-1	4.13E-1	3.49E-1	3.02E-1	2.38E-1	1.67E-1	1.04E-1	5.39E-2	2.0
		6.19E-1	5.67E-1	5.23E-1	4.84E-1	4.51E-1	3.85E-1	3.35E-1	2.66E-1	1.88E-1	1.19E-1	6.17E-2	2.5
		6.51E-1	6.00E-1	5.56E-1	5.18E-1	4.84E-1	4.16E-1	3.64E-1	2.92E-1	2.08E-1	1.32E-1	6.90E-2	3.0
		6.78E-1	6.28E-1	5.85E-1	5.47E-1	5.13E-1	4.44E-1	3.91E-1	3.15E-1	2.26E-1	1.45E-1	7.60E-2	3.5
		7.01E-1	6.52E-1	6.10E-1	5.72E-1	5.38E-1	4.68E-1	4.14E-1	3.36E-1	2.43E-1	1.57E-1	8.28E-2	4.0
		7.20E-1	6.74E-1	6.32E-1	5.94E-1	5.61E-1	4.91E-1	4.36E-1	3.56E-1	2.59E-1	1.68E-1	8.93E-2	4.5
		7.38E-1	6.92E-1	6.51E-1	6.14E-1	5.81E-1	5.11E-1	4.56E-1	3.74E-1	2.75E-1	1.79E-1	9.56E-2	5.
		7.66E-1	7.23E-1	6.84E-1	6.49E-1	6.16E-1	5.47E-1	4.91E-1	4.07E-1	3.03E-1	2.00E-1	1.08E-1	6.
		[redacted]	7.49E-1	7.11E-1	6.77E-1	6.46E-1	5.78E-1	5.22E-1	4.36E-1	3.28E-1	2.19E-1	1.19E-1	7.
	6.	2.34E-1	[redacted]	7.34E-1	7.01E-1	6.71E-1	6.04E-1	5.49E-1	4.63E-1	3.52E-1	2.37E-1	1.30E-1	8.
	7.	2.11E-1	2.51E-1	[redacted]	7.22E-1	6.92E-1	6.27E-1	5.73E-1	4.87E-1	3.73E-1	2.54E-1	1.41E-1	9.
	8.	1.92E-1	2.30E-1	2.66E-1	[redacted]	7.11E-1	6.48E-1	5.94E-1	5.08E-1	3.93E-1	2.70E-1	1.51E-1	10.
	9.	1.77E-1	2.13E-1	2.47E-1	2.78E-1	[redacted]	6.90E-1	6.39E-1	5.55E-1	4.38E-1	3.07E-1	1.75E-1	12.5
$ for 2.5th %iles	10.	1.63E-1	1.98E-1	2.30E-1	2.60E-1	2.89E-1	[redacted]	6.75E-1	5.93E-1	4.76E-1	3.40E-1	1.97E-1	15.
	12.5	1.38E-1	1.68E-1	1.97E-1	2.24E-1	2.50E-1	3.10E-1	[redacted]	6.52E-1	5.38E-1	3.96E-1	2.38E-1	20.
	15.	1.19E-1	1.46E-1	1.72E-1	1.97E-1	2.21E-1	2.76E-1	3.25E-1	[redacted]	6.25E-1	4.83E-1	3.07E-1	30.
	20.	9.36E-2	1.16E-1	1.38E-1	1.59E-1	1.79E-1	2.28E-1	2.72E-1	3.48E-1	[redacted]	5.97E-1	4.10E-1	50.
	30.	6.56E-2	8.19E-2	9.83E-2	1.14E-1	1.30E-1	1.69E-1	2.05E-1	2.70E-1	3.75E-1	[redacted]	5.69E-1	100.
	50.	4.11E-2	5.18E-2	6.26E-2	7.35E-2	8.44E-2	1.11E-1	1.38E-1	1.87E-1	2.73E-1	4.03E-1	[redacted]	
	100.	2.13E-2	2.70E-2	3.28E-2	3.88E-2	4.49E-2	6.02E-2	7.56E-2	1.06E-1	1.63E-1	2.60E-1	4.31E-1	
		6.	7.	8.	9.	10.	12.5	15.	20.	30.	50.	100.	
		″ for 2.5th percentiles											

″ for 97.5th %iles

Table C.7. Acceptance limits for the Kolmogorov test of goodness of fit.

Sample size (n)	Significance level				
	0.20	0.15	0.10	0.05	0.01
1	0.900	0.925	0.950	0.975	0.995
2	0.684	0.726	0.776	0.842	0.929
3	0.565	0.596	0.636	0.708	0.829
4	0.493	0.525	0.565	0.624	0.734
5	0.447	0.474	0.509	0.563	0.669
6	0.410	0.435	0.468	0.519	0.617
7	0.381	0.405	0.436	0.483	0.576
8	0.358	0.381	0.410	0.454	0.542
9	0.339	0.360	0.387	0.430	0.513
10	0.323	0.343	0.369	0.409	0.489
11	0.308	0.327	0.352	0.391	0.468
12	0.296	0.314	0.338	0.375	0.449
13	0.285	0.302	0.325	0.361	0.432
14	0.275	0.292	0.314	0.349	0.418
15	0.266	0.282	0.304	0.338	0.404
16	0.258	0.274	0.295	0.327	0.392
17	0.250	0.266	0.286	0.318	0.381
18	0.244	0.259	0.279	0.309	0.371
19	0.237	0.252	0.271	0.301	0.361
20	0.232	0.246	0.265	0.294	0.352
25	0.208	0.221	0.238	0.264	0.317
30	0.190	0.202	0.218	0.242	0.290
35	0.177	0.187	0.202	0.224	0.269
40	0.165	0.176	0.189	0.210	0.252
50	0.148	0.158	0.170	0.188	0.226
60	0.136	0.144	0.155	0.172	0.207
70	0.126	0.134	0.144	0.160	0.192
80	0.118	0.125	0.135	0.150	0.179
Large n	$1.07/\sqrt{n}$	$1.14/\sqrt{n}$	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$	$1.63/\sqrt{n}$

Reject the hypothesized distribution $F(x)$ if $D = \max|F_n(x) - F(x)|$ exceeds the tabulated value.

The asymptotic formula gives values that are slightly too high — by 1% to 2% for $n = 80$.

Table C.8. Parameters of constrained noninformative prior for binomial *p*.

p_0	b	α
0.50	0.	0.5000
0.40	!0.8166	0.4168
0.30	!1.748	0.3590
0.20	!3.031	0.3243
0.15	!4.027	0.3211
0.10	!5.743	0.3424
0.09	!6.295	0.3522
0.08	!6.958	0.3648
0.07	!7.821	0.3802
0.06	!8.978	0.3980
0.05	!10.61	0.4171
0.04	!13.08	0.4358
0.03	!17.22	0.4531
0.02	!25.53	0.4693
0.01	!50.52	0.4848
0.005	!100.5	0.4925
0.001	!500.5	0.4985
0	!4	0.5000

The table gives parameters of the constrained noninformative prior for a binomial parameter p, when the assumed prior mean is p_0. The exact constrained noninformative prior has the form

$$f_{\text{prior}}(p) \propto e^{bp} p^{-1/2}(1-p)^{-1/2},$$

with b tabulated above.

The tabulated value α is the first parameter of a beta distribution that has the same mean and variance as the constrained noninformative prior. The second parameter of that beta distribution is obtained by solving $\alpha/(\alpha+\beta) = p_0$. This results in the formula $\beta = \alpha(1-p_0)/p_0$. Then a beta($\alpha$, β) distribution with these parameters approximates the exact constrained noninformative prior.

If $p_0 > 0.50$, the exact constrained noninformative prior has positive b. For example, if $p_0 = 0.70$, look in the table at the row for $1 - 0.70$, and see that b there is !1.748. Therefore, the exact constrained noninformative prior is

$$f_{\text{prior}}(p) \propto e^{bp} p^{-1/2}(1-p)^{-1/2},$$

with $b = +1.748$.

If $p_0 > 0.50$, the beta approximation is obtained by interchanging the roles of p_0 and $1 - p_0$ and of α and β in the above formulas. For example, if p_0 were 0.30, the beta approximation would have $\alpha = 0.3590$ and $\beta = 0.3590 \times 0.70/0.30 = 0.8377$. If instead $p_0 = 0.70$, the beta approximation has $\alpha = 0.8377$ and $\beta = 0.3590$.

D. GLOSSARY

Arithmetic mean. See **mean**.

Bathtub curve. A plot of the failure rate as a function of system age, showing a high failure rate when the system is new (the **burn-in** period), then a low, roughly constant failure rate, followed by a rising failure rate as the system becomes old (the **wear-out** period). See **burn-in failure**.

Bayesian inference. Statistical inference involving the use of Bayesian methods. Bayesian inference uses probability distributions to model uncertainty in unknown quantities. Thus, unknown parameters are treated formally as if they were random variables. See also **frequentist inference** and **statistical inference**.

Bias. The difference between the expected value of an estimator and the true quantity being estimated. For example, if Y is a function of the data that estimates a parameter θ, the bias of Y is $E(Y) - \theta$.

Bin. A group of values of a continuous variable, used to partition the data into subsets. For example, event dates can be grouped so that each year is one bin, and all the events during a single year form a subset of the data.

Burn-in failure. Failures associated with the early time-frame of a component's life cycle, during which the failure rate often starts from a maximum value and decreases rapidly. The high failure rate early in the component's life cycle can be caused by poor quality control practices and a natural wear-in or debugging period. See **bathtub curve**.

c.d.f. See **cumulative distribution function**.

Cell. When the data are expressed in a table of counts, a cell is the smallest element of the table. Each cell has an observed count and, under some null hypothesis, an expected count. Each cell can be analyzed on its own, and then compared to the other cells to see if the data show trends, patterns, or other forms of nonhomogeneity. In a $1 \times J$ table, as with events in time, each cell corresponds to one subset of the data. In a $2 \times J$ table, as with failures on demand, each data subset corresponds to two cells, one cell for failures and one for successes.

Central moment. See **moment**.

Coefficient of variation. See **relative standard deviation**.

Common-cause failure. A single event that causes failure of two or more components at the same time (also referred to as common-mode failure).

Confidence interval. In the frequentist approach, a $100p\%$ confidence interval has a probability p of containing the true unknown parameter. This is a property of the procedure, not of any one particular interval. Any one interval either does or does not contain the true parameter. However, any random data set leads to a confidence interval, and $100p\%$ of these contain the true parameter. Compare with **credible interval**.

Conjugate. A family of prior distributions is conjugate, for data from a specified distribution, if a prior distribution in the family results in the posterior distribution also being in the family. A prior distribution in the conjugate family is called a conjugate prior. For example, the gamma distributions are conjugate for Poisson data, and the beta distributions are conjugate for binomial data.

Credible interval. In the Bayesian approach, a $100p\%$ credible interval contains $100p\%$ of the Bayesian probability distribution. For example, if θ has been estimated by a posterior distribution, the 5th and 95th percentiles of this distribution contain 90% of the probability, so they form a (posterior) 90% credible interval. It is not required to have equal probability in the two tails (5% in this example), although it is very common. For example, the interval bounded by 0 and the 90th percentile would also be a 90% credible interval, a one-sided interval. Bayes credible intervals have the same intuitive purpose as frequentist confidence intervals, but their definitions and interpretations are different.

Cumulative distribution function (c.d.f.). This function gives the probability that the random variable does not exceed a given value x. For a random variable X, the c.d.f. $F(x) = \Pr(X \# x)$. If X is discrete, such as a count of events, the c.d.f. is a step function, with a jump at each possible value of X. If X is continuous, such as a duration time, the c.d.f. is continuous. See also **probability density function**. Do not confuse the statistics acronym c.d.f. with the PRA acronym CDF, denoting core damage frequency!

Density. See **probability density function**.

Disjoint. See **mutually exclusive**.

Duration. The time until something of interest happens. The thing of interest may be failure to run, recovery from a failure, restoration of offsite power, etc.

Error factor. A representation of one of the parameters of the lognormal distribution, defined as the 95th percentile divided by the median. The error factor is a measure of the spread of the distribution, and is denoted by EF.

Estimate, estimator. In the frequentist approach, an **estimator** is a function of random data, and an **estimate** is the particular value taken by the estimator for a particular data set. That is, the term **estimator** is used for the random variable, and **estimate** is used for a number. The usual convention of using upper-case letters for random variables and lower-case letters for numbers is often ignored in this setting, so the context must be used to show whether a random variable or a number is being discussed.

Event rate. See **failure rate** for repairable systems, and replace the word "failure" by "event."

Expected value. If X is discrete with p.d.f. f, the expected value of X, denoted $E(X)$, is $\sum x_i f(x_i)$. If instead X is continuously distributed with density f, the expected value is $\int x f(x) dx$. The expected value of X is also called the **mean** of X. It is a measure of the center of the distribution of X.

Exposure time. The length of time during which the events of interest can possibly occur. The units must be specified, such as reactor-critical-years, site-calendar-hours, or system-operating-hours. Also called **time at risk**.

Failure on demand. Failure when a standby system is demanded, even though the system was apparently ready to function just before the demand. It is modeled as a random event, having some probability, but unpredictable on any one specific demand. Compare **standby failure**.

Failure rate. For a repairable system, the failure rate, θ, is such that $\theta \Delta t$ is approximately the expected number of failures in a short time period from t to $t + \Delta t$. If simultaneous failures do not occur, $\theta \Delta t$ is also approximately the probability that a failure will occur in the period from t to $t + \Delta t$. In this setting, θ is also called a **failure frequency**. For a nonrepairable system, $\theta \Delta t$ is approximately the probability that an unfailed system at time t will fail in the time period from t to $t + \Delta t$. In this setting, θ is also called the **hazard rate**.

Fractile. See **quantile**.

Frequency. For a repairable system, **frequency** and **rate** are two words with the same meaning, and are used interchangeably. If simultaneous events do not occur, the frequency $\theta(t)$ satisfies $\theta(t) \Delta t$. Pr (an event occurs between t and $t + \Delta t$), for small Δt.

Frequentist inference. Statistical inference that interprets the probability of an event as the long-term relative frequency of occurrence of the event, in many repetitions of an experiment when the event may or may not occur. Unknown parameters are regarded as fixed numbers, not random. See also **Bayesian inference** and **statistical inference**.

Geometric mean. The geometric mean is an estimator of the location or center of a distribution. It is applicable only for positive data. The geometric mean, say $\tilde{t}$, for $t_1, t_2, \ldots, t_n$, is defined as

$$\tilde{t} = \exp\left[(1/n)\Sigma \ln t_i\right].$$

It is always less than or equal to the arithmetic mean.

Goodness of fit. This term refers to a class of nonparametric methods that are used to study whether or not a given set of data follows a hypothesized distribution. Both hypothesis tests and graphical methods are used to investigate goodness of fit.

Hazard rate. For a nonrepairable system, **hazard rate** and **failure rate** are two phrases with the same meaning, used interchangeably. The hazard rate $h(t)$ satisfies $h(t) \Delta t$. $\Pr(t < T \le t + \Delta t \mid T > t)$, where Δt is small and T denotes the duration time of interest. The hazard rate is also called the **hazard function**.

Hypothesis. A statement about the model that generated the data. If the evidence against the null hypothesis, H_0, is strong, H_0 is rejected in favor of the alternative hypothesis, H_1. If the evidence against H_0 is not strong, H_0 is "accepted"; that is, it is not necessarily believed, but it is given the benefit of the doubt and is not rejected.

Improper distribution. A function that is treated as a probability distribution function (p.d.f.), but which is not a p.d.f. because it does not have a finite integral. For example, a uniform distribution (constant p.d.f.) on an infinite range is improper. Improper distributions are sometimes useful prior distributions, as long as the resulting posterior distribution is a proper distribution.

Independent. See **statistical independence**.

Inference. See **statistical inference**.

Initiating event. Any event, either internal or external to the plant, that triggers a sequence of events that challenge plant control and safety systems, whose failure could potentially lead to core damage or large early release.

Interval. The notation (a, b) denotes the interval of all points from a to b. This is enough for all the applications in this handbook. However, sometimes an additional refinement is added, giving a degree of mathematical correctness that most readers may ignore: The standard notation in mathematics is that (a, b) includes the points between a and b, but not the two end points. In set notation, it is $\{x \mid a < x < b\}$. Square brackets show that the end points are included. Thus, $(a, b]$ includes b but not a, $\{x \mid a < x \# b\}$.

Interval estimate. One way of estimating a parameter is to identify that it falls in some interval (L, U) with a specified degree of certainty, or confidence. The interval (L, U) is referred to as an interval estimate of the parameter. L and U are calculated from the random data. The frequentist interval estimate is referred to as a confidence interval. It does not give a probability statement about the true parameter value. Rather, the interpretation of a $100(1 - \alpha)\%$ confidence interval is that, if the random data were drawn many times, $100(1 - \alpha)\%$ of the resulting interval estimates would contain the true value. A Bayesian interval estimate is referred to as a subjective probability interval, or credible interval, and can be interpreted as giving a subjective probability statement about the true parameter value being contained in the interval. Compare with **point estimate**. See also **confidence interval**, **credible interval**.

Inverse c.d.f. algorithm. An algorithm for generating random numbers (presented in Section 6.3.2.5.4).

Latin hypercube sampling (LHS). See **Monte Carlo simulation**.

Likelihood. For discrete data, the likelihood is the probability of the observations. For continuous data, the likelihood is the joint density of the observations, which is the product of the densities of the individual observations if the observations are independent. When some of the observations are discrete and some are continuous, the likelihood is the product of the two types. The likelihood is typically treated as a function of the parameters, with the data regarded as fixed.

Markov Chain Monte Carlo (MCMC). See **Monte Carlo simulation**.

Maximum likelihood estimator. For data generated from a distribution with one unknown parameter, say θ, the maximum likelihood estimate (MLE) of θ is the parameter value that maximizes the likelihood of the data. It is a function of the data, and is commonly denoted $\hat{\theta}$. The MLE is a popular frequentist estimator for two reasons. (1) In commonly used models, the MLE is an intuitively natural function of the data. (2) Under certain, commonly valid, conditions, as the number of observations becomes large the MLE is approximately unbiased with approximately the minimum possible variance, and is approximately normally distributed.

Mean. The mean, μ, of a random variable X is the weighted average of the outcomes, where the weights are the probabilities of the outcomes. More precisely, the mean of X is the expected value $E(X)$, $\sum x_i f(x_i)$ if X is discrete with p.d.f. f, and $\int x f(x) dx$ if X is continuously distributed with density f. See also **expected value**.

Mean square error or **mean squared error**. The expected squared difference between an estimator and the true quantity being estimated. For example, if Y is a function of the data that estimates a parameter θ, the mean squared error (MSE) of Y is $E[(Y - \theta)^2]$. It can be shown that the $\text{MSE}(Y) = \text{var}(Y) + [\text{bias}(Y)]^2$.

Median. For a random variable X with a continuous distribution, the median is that value m for which $\Pr(X < m) = 0.5$, and thus also $\Pr(X > m) = 0.5$. For a sample of data values, or for a discrete random variable X taking a finite number of values with equal probability, the median is the middle value in the ordered set of values. The median m is the 50th percentile, $x_{0.50}$. See **percentile** for the general definition.

Mode. A mode of a distribution is a local maximum value of the probability density or probability distribution function (p.d.f.). A normal distribution has a single mode, which measures the center of the distribution.

Moment. The *k*th moment about *a* of a random variable *X* is the expected value of $(X - a)^k$. If *X* is discrete with p.d.f. *f*, this is $\sum (x_i - a)^k f(x_i)$. If *X* is continuous with density *f*, the *k*th moment about *a* is $\int (x - a)^k f(x)dx$. The moments about 0 are sometimes called simply "the" moments. Moments about the mean are called **central moments**. The first moment is the mean, often denoted μ. The second central moment is the variance.

Monte Carlo Sampling. See **Monte Carlo simulation**.

Monte Carlo simulation. Generally referred to as Monte Carlo Sampling by probabilistic risk assessment (PRA) analysts, Monte Carlo simulation uses a sample of values from an uncertainty distribution to approximate the distribution. Moments and percentiles of the distribution are approximated by the corresponding moments and percentiles of the sample.

In the usual PRA application, the initiating event frequencies and basic event probabilities have Bayesian distributions. They are sampled by simple random sampling or by Latin hypercube sampling (LHS). The parameter values are propagated through a fault-tree/event-tree model to produce a simulation of the uncertainty distribution of a quantity of interest, such as core damage frequency.

Monte Carlo simulation is also used to approximate the posterior distributions of parameters, when direct calculations are not feasible. In situations with a single parameter and a nonconjugate prior, a simple random sample from the posterior distribution can usually be constructed. In more complicated situations, such as those with multiple interrelated parameters, the unknown parameters can be simulated by Markov chain Monte Carlo (MCMC) to produce a sequence, or chain, of values of each parameter. The values are not independent. However, if the initial portion of the chain is discarded, the remainder of the terms mimic the posterior distribution of the parameter.

Mutually exclusive. Events are mutually exclusive, or **disjoint**, if no two of them have any elements in common. The intersection of any two of the events is the empty set.

Nonparametric. In parametric inference, the data are assumed to come from a known distributional form, with only the parameters unknown. In nonparametric inference, no distributional form is assumed. Not only are the values of the parameters unknown, but the form of the distribution is unknown as well. See **parametric**.

Nonrepairable system. A system that can only fail once, after which data collection stops. An example is a standby safety system, if the failure to run cannot be recovered during the mission of the system. Data from a nonrepairable system consist of data from identical copies of the system. For example, data from a safety system may be collected, with each run starting with the system nominally operable, and the system either running to completion of the mission or failing before that time. The successive demands to run are regarded as demands on identical copies of the system. See **repairable system**.

Null hypothesis. See **hypothesis**.

Order statistics. The random values arranged from smallest to largest. For example, suppose that three times are observed, with $t_1 = 8.4$, $t_2 = 3.0$, and $t_3 = 5.1$. The order statistics are $t_{(1)} = 3.0$, $t_{(2)} = 5.1$, and $t_{(3)} = 8.4$. Before the data are observed, one can consider the order statistics as random variables, $T_{(1)}, T_{(2)}, \ldots, T_{(n)}$.

Outage, outage time. An outage is an event when a system is unavailable, that is, out of service for some reason. The outage time is the duration of the event. Compare with **unavailability**.

Parameter. A parametric family of distributions is a collection of distributions that is indexed by one or more quantities called parameters. For example, suppose that $f(t; \theta) = \theta e^{-t\theta}$, where t; $\theta > 0$. For each value of θ, $f(t; \theta)$ is a probability density function. Here θ is the parameter that identifies the particular density in the family of exponential density functions. The normal family has two parameters, the mean and the variance.

Parametric. Parametric statistical inference is concerned with learning the values of unknown parameters (and their associated properties) from sample data for a given or assumed family of distributions. See **nonparametric**.

p.d.f. See **probability density function** and **probability distribution function**.

Percentile. Consider a continuous distribution with density (p.d.f.) *f* and cumulative distribution function (c.d.f.) *F*. The 100*q*th percentile is the value *x* such that $F(x) = q$, or equivalently

$$\int_{-\infty}^{x} f(u)du = q .$$

If the distribution is concentrated on the positive line, the lower limit of integration may be replaced by 0. The 100*q*th percentile is equal to the *q*th quantile. For example, the 95th percentile equals the 0.95 quantile. If *X* has a discrete distribution, a percentile may not be unique. The 100*q*th percentile is defined in this case as *x* such that Pr(*X* # *x*) $ 100*q*% and Pr(*X* $ *x*) $ 100(1 ! *q*)%.

Similarly, for a finite sample, the 100*q*th percentile is defined as *x* such that at least 100*q*% of the values in the sample are *x* or smaller, and at least 100(1 ! *q*)% are *x* or larger. For example, if a sample is a set of three numbers, {1.2, 2.5, 5.9}, the median (corresponding to $q = 0.5$) is 2.5, because at least half of the numbers are 2.5 or smaller and at least half are 2.5 or larger. If the sample has four numbers, {1.2, 2.5, 2.8, 5.9}, then any number from 2.5 to 2.8 can be considered a median. In this case, the average, (2.5 + 2.8)/2, is often chosen.

Point estimate. An estimate of a parameter in the form of a single number is called a point estimate of the parameter. For example, the mean of a sample of values of a random variable *X* is a commonly used point estimate of the mean of the distribution. Compare with **interval estimate**.

Poisson process. A process in which events (such as failures) occur in a way such that the number of events *X* in total time *t* is described by a Poisson distribution. See Section 2.2.2, Section 7.2, or Appendix A.6.2 for more details.

Pool. To combine data from distinct sources, ignoring possible differences between the sources. Data are sometimes pooled from distinct time periods, components, trains, and/or power plants.

Population. In the PRA setting, population refers to the random distribution that generates data. Population attributes, such as the population mean or population median, are those attributes of the probability distribution. Compare with **sample**.

Posterior credible interval. See **credible interval**.

Posterior distribution. A distribution that quantifies, in a Bayesian way, the belief about a parameter after data have been observed. It reflects both the prior belief and the observed data.

Power of a test. The probability that the test will reject H_0 when H_0 is false. If many possible alternatives to H_0 are considered, the power depends on the particular alternative. See **hypothesis**.

Prior. A colloquial abbreviation for **prior distribution**.

Prior distribution. A distribution that quantifies, in a Bayesian way, the belief about a parameter before any data have been observed.

Probability model. A term for the set of mathematical relationships which are used to define both cumulative distribution functions and either probability distribution functions (discrete case) or probability density functions (continuous case).

Probability density function (p.d.f.). For a continuous random variable *X*, the probability density function *f* satisfies

$$\Pr(a \le X \le b) = \int_a^b f(x)dx \quad .$$

Properties of the density are

$f(x)$ $ 0 for all x

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$f(x)$) x . Pr($x < X$ # x +) x) for small) x.

The p.d.f. is related to the c.d.f. by

$f(x) = F/(x)$, the derivative,

and

$$F(x) = \int_{-\infty}^{x} f(u)du \ .$$

See **cumulative distribution function**.

Probability distribution function (p.d.f.). For a discrete random variable *X*, the p.d.f. $f(x) = \Pr(X = x)$.

p-value. In the context of testing, the p-value is the significance level at which the data just barely cause H_0 to be rejected. H_0 is rejected when a test statistic is extreme, and the p-value is the probability (under H_0) that the random test statistic would be at least as extreme as actually observed.

Quantile. Consider a continuous distribution with density (p.d.f.) *f* and cumulative distribution function (c.d.f.) *F*. The *q*th quantile is the value *x* such that $F(x) = q$, or equivalently:

$$\int_{-\infty}^{x} f(u)du = q$$

If the distribution is concentrated on the positive line, the lower limit of integration may be replaced by 0. The qth quantile is equal to the $(100q)$th percentile. For example, the 0.95 quantile equals the 95th percentile. If X has a discrete distribution, a quantile may not be unique. Some authors use the term **fractile** instead of quantile. See **percentile** for a fuller explanation.

Random sample. $x_1, ..., x_n$ are a random sample if they are the observed values of $X_1, ..., X_n$, where the X_is are statistically independent of each other and all have the same distribution.

Random variable. A rule that assigns a number to every outcome in a sample space. For example, if a pump was demanded to start n times, the sample space consists of all the possible outcomes, with their probabilities. A random variable of interest might be the *number* of failures to start. If a stuck valve is repaired, the sample space consists of all the possible outcomes of the repair process, with their probabilities. A random variable of interest might be the time required for repair, a *number*.

Range. The difference between the largest and smallest values of a sample is called the range of the sample.

Rate. See **frequency**.

Reactor critical year. 8760 hours during which a reactor is critical.

Rejection-method algorithm. An algorithm for generating a random sample from a particular distribution. Its general form is given in Section 6.2.2.6, and applied in several places there and in Section 6.3.2.4.

Relative standard deviation. The standard deviation, expressed as a fraction of the mean. The relative standard deviation of X is st.dev.$(X)/E(X)$. Some authors call it the **coefficient of variation**, and express it as a percent.

Relative variance. The square of the **relative standard deviation**. The relative variance of X is $\text{var}(X)/[E(X)]^2$.

Renewal process. A process in which events (such as failures or restorations) occur in a way such that the times between events are independent and identically distributed. For example, if the process consists of failures and nearly instantaneous repairs, each repair restores the system to good-as-new condition.

Repairable system. A system that can fail repeatedly. Each failure is followed by repair, and the possibility of another failure sooner or later. An example is a power plant, with initiating events counted as the "failures." After such an event, the plant is brought back up to its operating condition, and more initiating events can eventually occur. See **nonrepairable system**.

Residual. When a model is fitted to data, the residual for a data point is the data value minus the fitted value (the estimated mean). The residuals together can be used to quantify the overall scatter of the data around the fitted model. If the assumed model assigns different variances to different data points, the standardized residuals are sometimes constructed. A **standardized residual** is the ordinary residual divided by its estimated standard deviation.

Return-to-service test. A test performed at the end of maintenance, which must be successful. If the system does not perform successfully on the test, the maintenance is resumed and the test is not counted as a return-to-service test. A return-to-service test can demonstrate that no latent failed conditions exist (see **standby failure**), but it provides absolutely no information about the probability of failure on a later demand (see **failure on demand**).

Sample. This term refers to data that are generated randomly from some distribution. Sample attributes, such as the sample mean or sample median, are those attributes calculated from the sample. They may be used as estimators of the corresponding population attributes. The sample may be thought of as random, before the data are generated, or as fixed, after the data are generated. See also **population**, **random sample**, **sample mean**, **sample median**, and **sample variance**.

Sample mean. The arithmetic average of the numbers in a random sample. If the numbers are $x_1, ... , x_n$, the sample mean is often denoted $\bar{x}$. It is an estimate of the **population mean**, that is, of the expected value $E(X)$.

Sample median. Let $x_{(1)}, ... x_{(n)}$ be the order statistics from a random sample. The sample median is the middle value. If n is odd, the sample median is $x_{([n+1]/2)}$. If n is even, the sample median is any number between $x_{(n/2)}$ and $x_{(n/2+1)}$, although usually the average of these two numbers is used.

Sample variance. Let $x_1, ... , x_n$ be a random sample, with sample mean $\bar{x}$. The sample variance, often denoted s^2, is

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \ .$$

It is an estimate of the **population variance**, var(X).

Significance level of a test. The probability of making a Type I error, that is, of rejecting H_0 when H_0 is true. (If H_0 includes a number of possibilities, so that the probability of rejecting H_0 varies, the significance level is defined as the maximum of those probabilities.) The significance level is denoted by α. Compare with **p-value** and **statistically significant**.

Skewed distribution. A distribution that is not symmetrical. A distribution that is restricted to the range from 0 to ∞ is typically skewed to the right, or positively skewed. Its mean is larger than its median, and the 95th percentile is farther from the median than the 5th percentile is. The Poisson, gamma, and lognormal distributions are a few examples of positively skewed distributions.

Standard deviation. The standard deviation of a distribution is the square root of the variance. The standard deviation and variance are two measures of how much spread or dispersion there is in a distribution.

Standard error. The estimated standard deviation of the estimator of a parameter, in the frequentist approach. For example, suppose that θ is the parameter to be estimated, and $\hat{\lambda}$ is the estimator. The estimator depends on random data, and therefore is random, with a standard deviation, s.d.($\hat{\lambda}$). The estimated value of this standard deviation is the standard error for θ.

Standardized residual. See **residual**.

Standby failure. For a standby system, failure to start resulting from an existing, or latent, failed condition. The system is in this failed condition for some time, but the condition is not discovered until the demand. Compare **failure on demand**.

Statistic. A function of the data, such as the sample mean or the Pearson chi-squared statistic. Before the data are observed, the statistic is a random variable which can take many values, depending on the random data. The observed value of a statistic is a number.

Statistical independence. Two events are statistically independent if the probability of both occurring is the product of their marginal (or individual) probabilities: $\Pr(E_1 \cap E_2) = \Pr(E_1) \times \Pr(E_2)$. Three or more events are statistically independent if the probability of any set of the events is equal to the product of the probabilities of those events. Two or more random variables are statistically independent if their joint p.d.f. equals the product of the marginal (or individual) p.d.f.s. For brevity, the word *statistically* is often dropped.

It can be shown that two random variables are statistically independent if and only if any event defined in terms of one random variable is statistically independent of any event defined in terms of the other random variable. (A similar statement holds for more than two random variables.) For example, suppose that X and Y are independent continuously distributed random variables, with joint density

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \ .$$

Let A be the event $a \le X \le b$, and let B be the event $c \le Y \le d$. Then

$$\Pr(A \cap B) = \Pr(a \le X \le b \text{ and } c \le Y \le d)$$

$$= \int_a^b \int_c^d f_{X,Y}(x,y)\,dy\,dx$$

by the definition of a joint density

$$= \int_a^b \int_c^d f_X(x) f_Y(y)\,dy\,dx$$

because X and Y are independent

$$= \int_a^b f_X(x)\,dx \int_c^d f_Y(y)\,dy$$

evaluating the integral

$$= \Pr(a \le X \le b)\Pr(c \le Y \le d)$$

by definition of the marginal densities

$$= \Pr(A) \times \Pr(B).$$

Statistical inference. The area of statistics concerned with using sample data to answer questions and make statements about the distribution of a random variable from which the sample data were obtained.

Statistically significant. A departure from a null hypothesis is called statistically significant if the hypothesis is rejected with some small significance level, customarily set to 0.05. See **p-value** and **significance level of a test**.

Stochastic. Referring to a random, rather than a deterministic, process. This is an elevated word for *random*.

System. In this handbook, system is the general word used to denote a collection of hardware for which data are collected. The term can apply to a specific system typically found in a nuclear power plant, such as the auxiliary feedwater system, or to a train, or a component, or even a small piece part, as long as data for the system are reported.

Time at risk. See **exposure time**.

Type I error. A rejection of the null hypothesis when it is true.

Type II error. "Acceptance" of the null hypothesis when it is false, that is, failure to reject the null hypothesis when it is false.

Unavailability. For a standby system, the probability that the system is unavailable, out of service, when demanded. This may be divided into different causes — unavailability from planned maintenance and unavailability from unplanned maintenance. Unavailability is distinct from failure to start of a nominally available system. Compare **outage**.

Uncertainty. The imprecisions in the analyst's knowledge or available information about the input parameters to PRA models, the PRA models themselves, and the outputs from such models.

Variance. The variance of a random variable X, denoted by σ^2, is the second moment about the mean, the average of the squared deviations from the mean, $E[(X - \mu)^2]$. It measures the dispersion in the distribution. Compare **standard deviation**.

INDEX

Page

Page

Page

Page

Page